THE
PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 71, No. 10 MAY 15, 1947

Range-Momentum Measurements of Particles Emitted in Nuclear Disintegrations Induced by 100-Mev X-Rays*

G. S. KLAIBER, E. A. LUEBKE, AND G. C. BALDWIN
Research Laboratory, General Electric Company, Schenectady, New York

(Received February 10, 1947)

The masses of the particles have been determined by measurement of magnetic rigidity and range of tracks originating from bombardment with 100-Mev x-rays of the glass wall of a cloud chamber. Range was determined by observing the stopping in a series of aluminum foils mounted in the cloud chamber and the magnetic rigidity was determined from measurement of curvature of the track before it entered the foils. The magnetic field of 2650 gauss permits particles of intermediate mass to be readily distinguished from protons. About 200 heavily ionizing particles were observed in a total of 3100 pictures. 102 of these tracks stopped either in the gas of the chamber or in one of the foils. Photographs of typical tracks are presented. None of these 102 tracks could be identified as caused by a particle having a mass in the neighborhood of, or less than, the meson mass (i.e., 200 electron masses). All of these tracks showed positive curvature. The majority could definitely be identified as protons. The few deviations can be accounted for by considering the effects of multiple scattering in the gas of the cloud chamber. Several particles were observed to penetrate all of the foils; protons with energies in excess of 28 Mev are indicated. The energy distribution of the protons is given.

IN previous work,[1,2] it has been shown that disintegrations of nuclei with the emission of heavy charged particles are induced by 100-Mev x-radiation. The present work was undertaken to establish more certainly the identity of the heavy particles emitted in these disintegrations.

The Wilson cloud chamber affords a convenient means of determining the mass of these particles. There are four properties of a particle, namely, the range, magnetic rigidity ($H\rho$), density of ionization, and rate of change of $H\rho$ with range, which can be determined with the cloud chamber. As pointed out by Corson and Brode,[3] any two of these quantities are sufficient to determine the mass of a particle, but of the various combinations of these four quantities, the use of range and magnetic rigidity affords the most accurate determination of the mass. These are the two quantities which were used for the mass determinations in this work. The ranges of the particles were determined by observing their stopping in a series of foils mounted in the cloud chamber. The magnetic rigidity was determined from a measurement of the curvature of that part of the track generated by the particle before it entered the stopping foils.

A schematic drawing of the experimental set-up is shown in Fig. 1. The x-ray beam from the 100-Mev betatron[4] is collimated by an aperture $1\frac{1}{4}$ inches high and $1\frac{3}{4}$ inches wide in a lead

* Presented in part at the September, 1946 meeting of the American Physical Society in New York City.

[1] G. C. Baldwin and G. S. Klaiber, Phys. Rev. 70, 259 (1946).

[2] M. Schein, A. J. Hartzler, and G. S. Klaiber, Phys. Rev. 70, 435 (1946).

[3] D. R. Corson and R. B. Brode, Phys. Rev. 53, 773 (1938).

[4] W. F. Westendorp and E. E. Charlton, J. App. Phys. 16, 581 (1945).

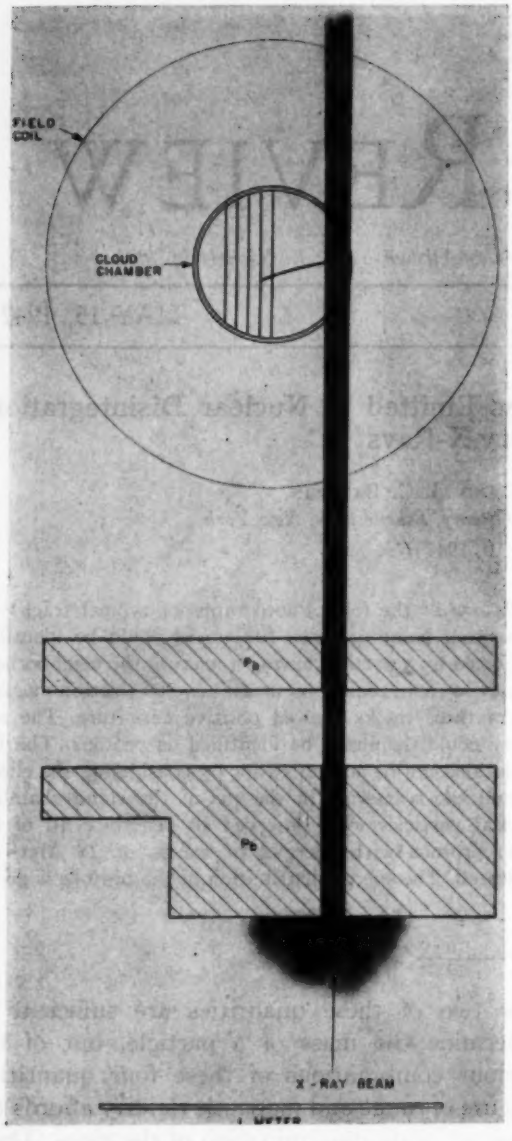


FIG. 1. Schematic diagram of experimental arrangement.

shield having a total thickness of 16 inches. The collimated beam of x-rays is then allowed to strike one edge of the glass wall of the cloud chamber. Heavy particles emitted by the photodisintegrations in the glass can then be observed traveling across the chamber, first through an unobstructed part of the chamber where curvature measurements can readily be made, and finally into the series of aluminum foils in which they are stopped.

The cloud chamber used in this work had a diameter of 30 cm and a depth of 3 cm of the chamber was illuminated. The source of illumination was two xenon capillary lamps mounted on opposite sides of the chamber. The aluminum foils in the chamber increased in thickness in logarithmic steps from 0.006″ for the first foil to 0.080″ for the last. The chamber was filled with air or hydrogen and a mixture of alcohol and water vapors to a pressure of one atmosphere. A pair of Helmholtz coils supplied a magnetic field of 2650 gauss perpendicular to the plane of the chamber. The chamber was operated once a minute in synchronism with the betatron.

The alignment of the hole in the lead shield with the very narrow 100-Mev x-ray beam was accomplished using radiographic techniques. An 8-×10-inch x-ray film wrapped in thin black paper was placed against the side of the collimator shield facing the betatron and in such a position that the aperture of the collimator was at the center of the film. A $\frac{1}{4}$-inch sheet of lead was then placed in front of the film to act as an intensifier. Figure 2 shows the radiograph obtained. Here one can see the circular spot due to the x-ray beam, the outline of the aperture, and their position relative to one another. The accuracy of the line-up used in this work is apparent from this picture.

The degree of collimation obtained with the 16 inches of lead used for this purpose can be

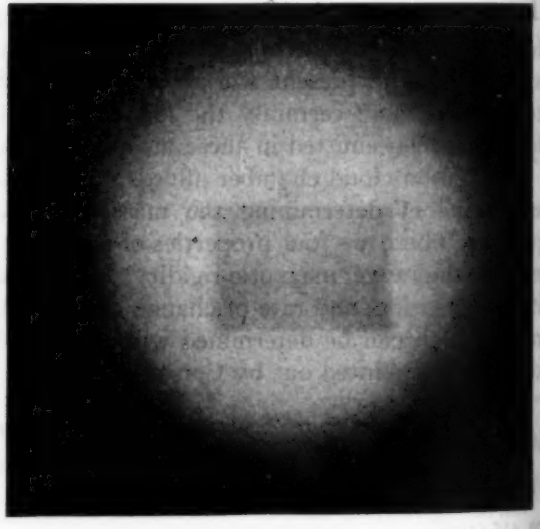


FIG. 2. Radiograph of 100-Mev x-ray beam incident on lead shield aperture.

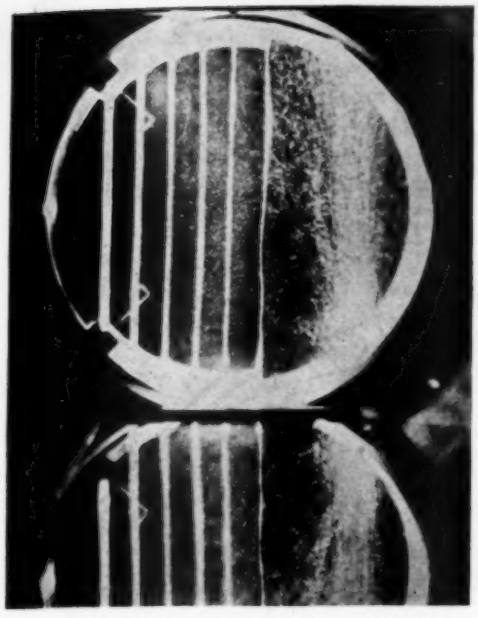


FIG. 3. Cloud-chamber photograph taken without magnetic field showing degree of collimation of x-ray beam.

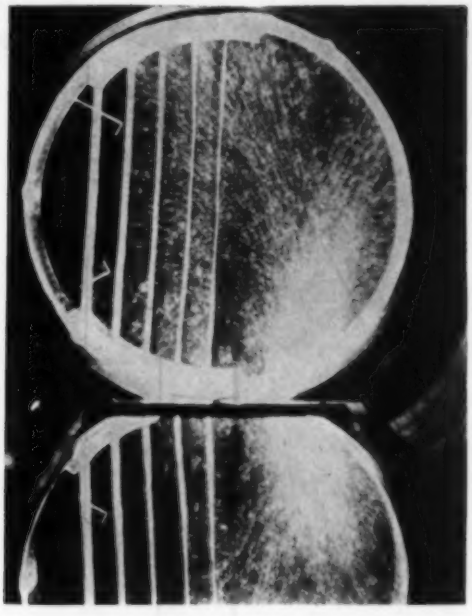


FIG. 4. Cloud-chamber photograph, similar to Fig. 3, taken with magnetic field of 2650 gauss.

seen in Fig. 3. For this picture, the chamber was placed so that the x-ray beam passed through the chamber instead of just hitting the edge. This picture was taken without a magnetic field. Since at high quantum energies, the pair and compton electrons produced in the chamber wall travel predominantly in a forward direction, they give a good indication of the extent of the x-ray beam. Figure 4 is similar to Fig. 3 except for a magnetic field of 2650 gauss applied perpendicular to the plane of the chamber. Here one can see very clearly the separation of the electrons and positrons in the electron beam from the glass wall. These pictures indicate that quite good collimation of the x-ray beam was achieved.

In selecting the thicknesses of the stopping foils used in this experiment, one criterion had to be satisfied. The foils had to be of such a thickness that for a particle stopping in any given foil, the uncertainty in its range, caused by the thickness of the foil, must be small enough compared to the total range of the particle to allow sufficient resolution in its mass determination to distinguish between protons and particles having masses of the order of the meson mass. That the foils used in this experiment satisfy this criterion can be seen in the following illustrations. Figure 5 is a diagram illustrating how a proton, a meson of mass 200 times the electron mass, and a particle of 50 times the electron mass should appear in the chamber with a magnetic field of 2650 gauss, if they stop in the first foil. The shaded area for each particle indicates the possible variations in the radius of curvature of that particle because of the uncertainty of its range owing to foil thickness. It can be seen that for particles stopping in the first foil, there should be no uncertainty as to the identity of the particle. Figure 6 is a diagram similar to Fig. 5 showing the resolution for particles stopping in the last foil. Here, as for particles stopping in the first foil, the resolution is still good. In a similar manner, it can be shown that for a particle stopping in any of the other foils, there is sufficient resolution in the mass determination to identify the particle.

A general summary of the resolution in the mass determination obtained in this experiment is shown in Fig. 7. Here each heavy diagonal line is a plot of range against magnetic rigidity for the particular particle indicated on the curve. The portions of the ranges corresponding to the various foils are indicated by the horizontal shaded areas on the graph. It can be seen from this diagram that for particles stopping in any foil there is no overlapping in the values of $H\rho$

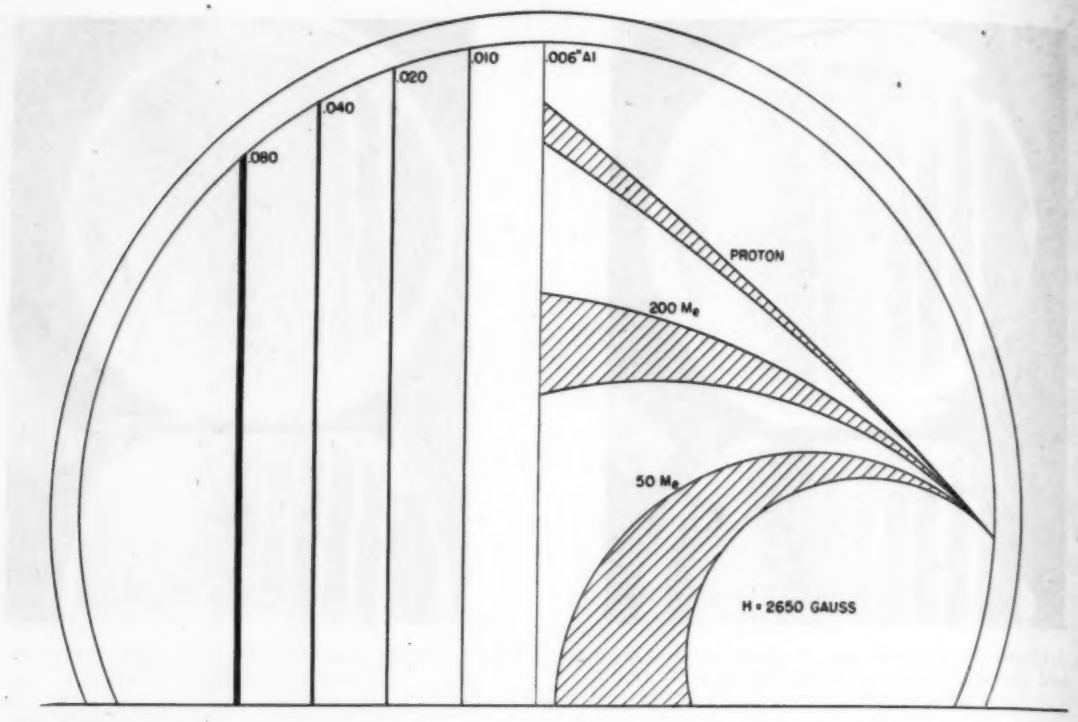


FIG. 5. Diagram illustrating how particles of mass indicated would appear in the cloud chamber with a magnetic field of 2650 gauss, if they stop in the first foil. The extent of the shaded area shows the variation of curvature with range due to foil thickness.

for any of the particles shown. It should also be noted that a meson of mass 200 with a range corresponding to stopping in the last foil has a smaller ρ than a proton stopping in the first foil.

A total of 3100 pictures were taken with the foil arrangement described above and with a magnetic field of 2650 gauss perpendicular to the plane of the chamber. There were about 200 heavily ionizing particles observed in this series of pictures. 102 of these observed tracks stopped either in the gas of the chamber or in one of the foils. Examples of some of these tracks are shown in the accompanying figures.

Figure 8 shows a particle having a range of 14.5 cm in air and a measured radius of curvature of 75 cm. These values of range and curvature are consistent for a particle of protonic mass having an energy of 3.1 Mev. A meson of mass 200 times the electron mass and having this range of 14.5 cm would generate a track having a radius of curvature of only 16 cm.

Figure 9 is of a particle stopping in the first foil. The range of this particle is between 13.5 and 44 cm of air (i.e., the equivalent thickness of the foil). The radius of curvature of this track is 100 cm. From Fig. 7, it can be seen that these data are consistent for a proton. A 200-mass meson with the extreme range of 44 cm would produce a track of only 26-cm radius of curvature. This particle is a proton having an energy somewhere in the range of 3 to 5.8 Mev.

In Fig. 10 a particle is shown stopping in the second foil. This particle has a range between 47 and 92 cm of air and its track exhibits a curvature with a radius of 120 cm. These data identify this particle to be a proton. The track of a meson of mass 200 having a range of 92 cm would have a radius of curvature of only 32 cm. This proton has an energy between 6 and 8.8 Mev.

Figure 11 shows a particle with a range somewhere between 105 and 195 cm of air and whose track has a radius of curvature of 200 cm. This identifies the particle as a proton with an energy between 9.5 and 13.5 Mev. The track of a meson of mass 200 with a 195-cm range should have a radius of curvature of 40 cm.

A particle stopping in the fourth foil is shown

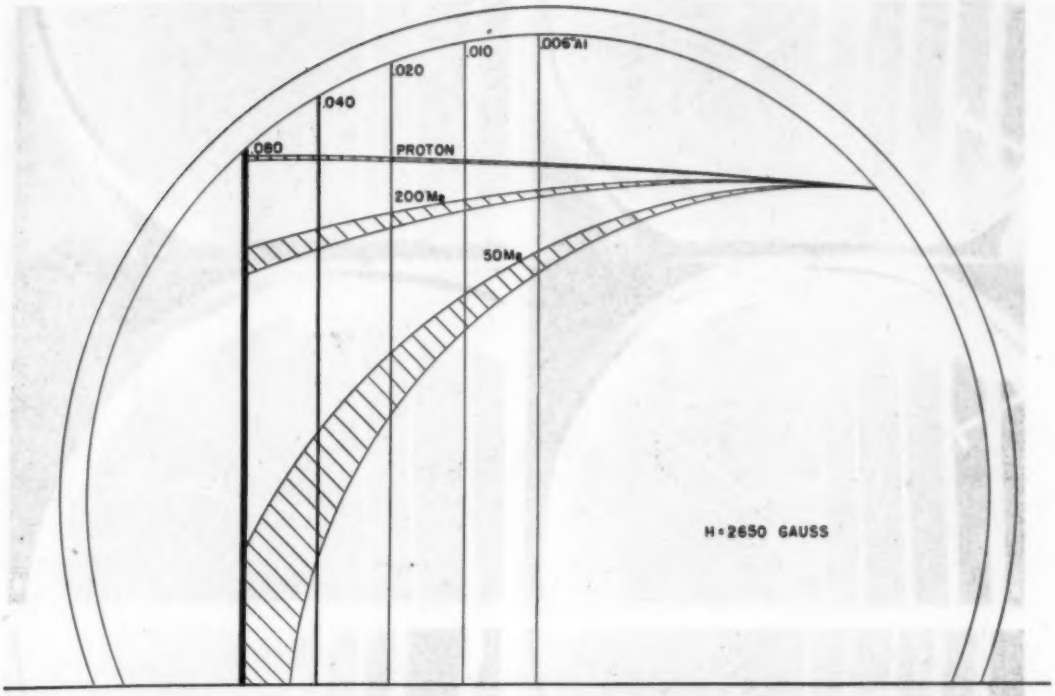


FIG. 6. Diagram illustrating how particles of mass indicated would appear in the cloud chamber with a magnetic field of 2650 gauss, if they stop in fifth plate. The extent of the shaded area shows the variation of curvature with range due to foil thickness.

in Fig. 12. The range of this particle is between 190 and 380 cm of air, and the radius of curvature of its track is about 200 cm. A 200-mass meson with a range of 380 cm would generate a track with only a 49-cm radius of curvature. This particle is a proton with an energy between 13.2 and 19.5 Mev.

Figure 13 shows a particle stopping in the fifth foil and thus having a range between 400 and 800 cm of air. The curvature of this track is so slight that it was impossible to get an exact measurement of its radius of curvature with the measuring device used for this purpose. The best that can be said is that the radius of curvature is greater than 250 cm. From Fig. 7 it can be seen that this particle can only be a proton. A meson of mass 200 with a range of 800 cm would have a track curvature with only a 60 cm radius. This proton has an energy between 20 and 30 Mev.

The examples that have been shown are typical of the 102 tracks that were observed to stop in the chamber. None of these 102 tracks could be identified as being caused by a particle having a mass in the neighborhood of, or less than, the meson mass (i.e., 200 times the electron mass). All of these tracks showed a positive curvature. While the majority of these 102 tracks

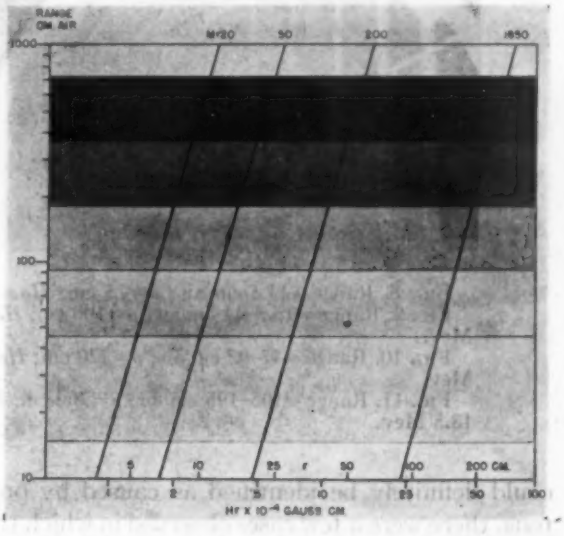


FIG. 7. Diagram summarizing resolution of mass determination as a function of foil thickness (horizontal shaded areas).

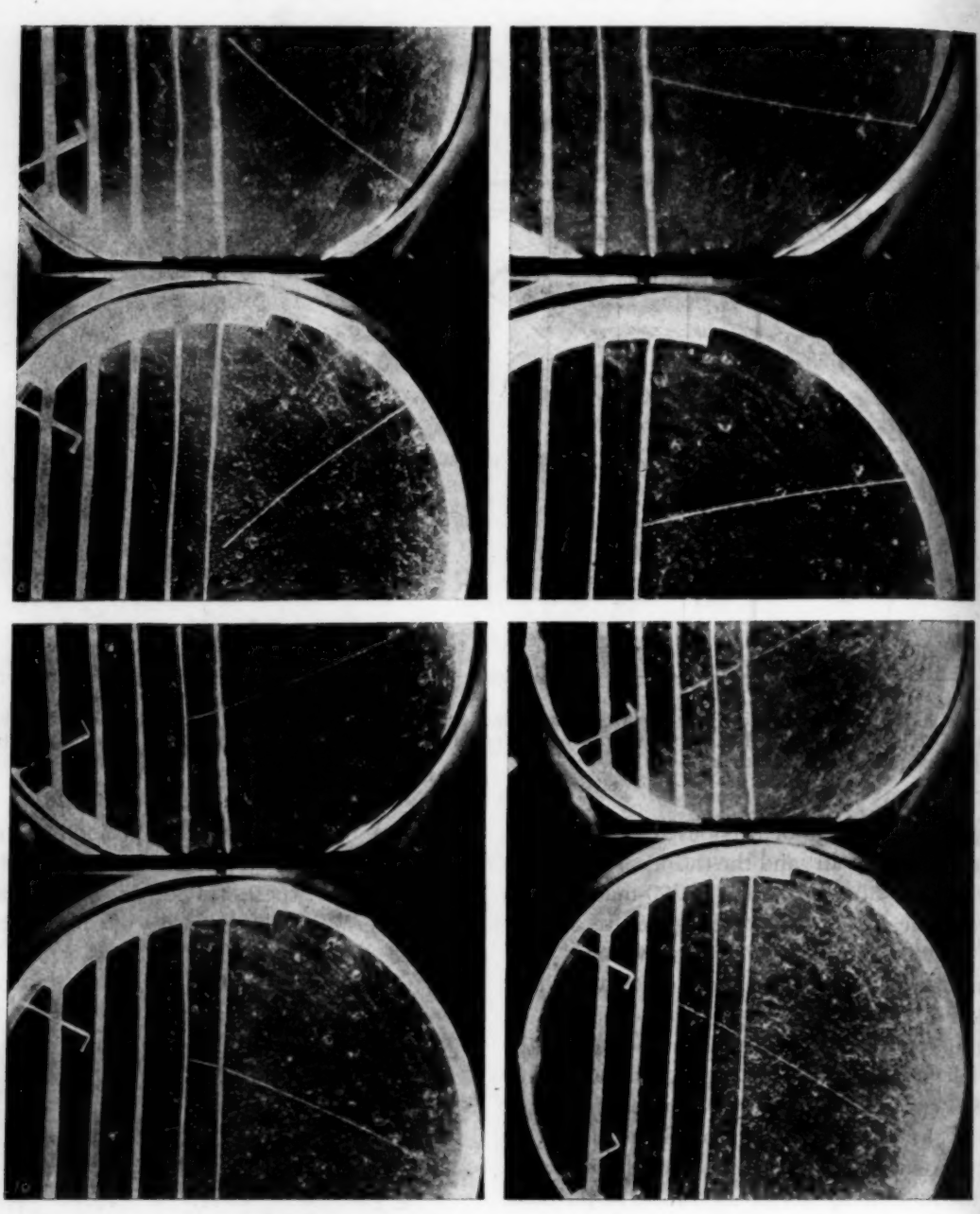


FIG. 8. Range = 14.5 cm air; $\rho = 75$ cm; $H\rho = 2 \times 10^5$ gauss cm; Identity: Proton, Energy = 3.1 Mev.
FIG. 9. Range = 13.5–44 cm air; $\rho = 100$ cm; $H\rho = 2.6 \times 10^5$ gauss cm; Identity: Proton, Energy = 3–5.8 Mev.
FIG. 10. Range = 47–92 cm air; $\rho = 120$ cm; $H\rho = 3.2 \times 10^5$ gauss cm; Identity: Proton, Energy = 6–8.8 Mev.
FIG. 11. Range = 105–195 cm air; $\rho = 200$ cm; $H\rho = 5.3 \times 10^5$ gauss cm; Identity: Proton, Energy = 9.5–13.5 Mev.

could definitely be identified as caused by protons, there were a few cases observed in which the mass of the particle, as calculated for its measured range and magnetic rigidity, deviated somewhat from the proton mass. In only two cases, however, did the mass determination lead to a value for the mass less than 1500 times the electron mass. One example of this is shown in

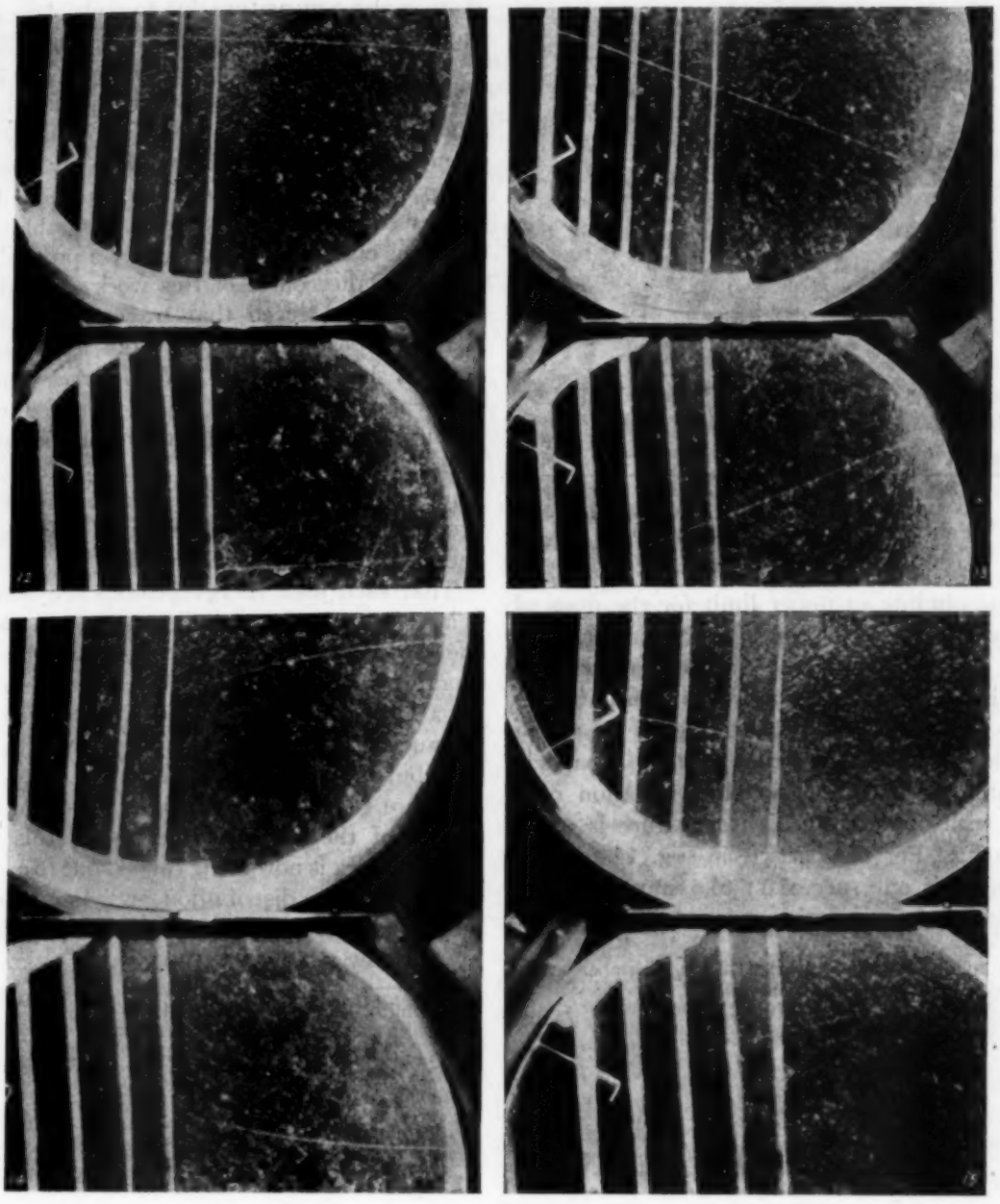


FIG. 12. Range = 190–380 cm air; $\rho \sim 200$ cm; $H\rho \sim 6.6 \times 10^5$ gauss cm; Identity: Proton, Energy = 13.2–19.5 Mev.
FIG. 13. Range = 400–800 cm air; $\rho > 250$ cm; $H\rho > 6.6 \times 10^5$ gauss cm; Identity: Proton, Energy = 20–30 Mev.
FIG. 14. Range > 14 cm air; $\rho = 45$ cm; $H\rho = 1.2 \times 10^5$ gauss cm; Mass < 1000 m_e; Identity: Proton if $\rho_{max} < 150$ cm.
FIG. 15. Range > 750 cm air; Energy (proton) > 28 Mev.

Fig. 14. This picture shows a particle whose range is greater than 14 cm of air. Only this lower limit can be given for the range as the particle runs out of the lighted portion of the chamber before stopping. The radius of curvature of this track is 45 cm, giving a magnetic rigidity for the particle of 1.2×10^5 gauss cm. Using these values of range and magnetic rigid-

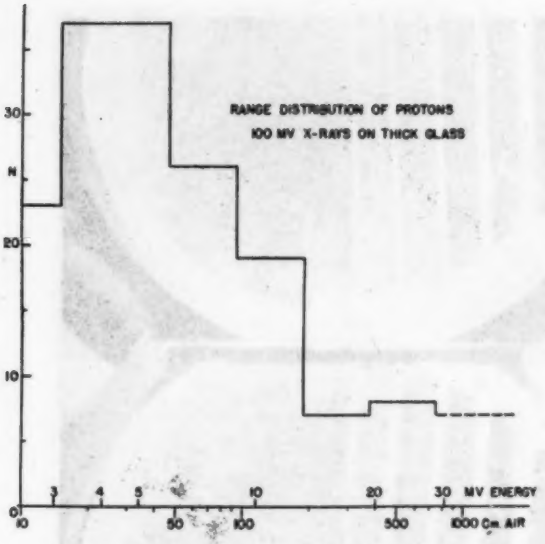


FIG. 16. Range distribution of protons.

ity, one obtains an upper limit for the mass of this particle of 1000 times the electron mass. While such a value for the mass appears offhand to be incompatable with a proton, consideration of the effect of multiple scattering by the gas in the cloud chamber on the path of a particle indicates that the occasional observation of a track with abnormally high curvature is to be expected.[5,6] For the particle in Fig. 14 to be a proton, it is only necessary to assume that, in addition to the magnetic curvature of the track, there is also a curvature due to multiple scattering, this scattered curvature having a radius of 150 cm. Such a curvature due to scattering is not unreasonable. A similar analysis of the few tracks that gave mass values deviating a little from the proton mass, showed in every case that the particle was most probably a proton. Thus, it was concluded that every measurable track that was observed was caused by a proton.

During the course of this work, several particles were observed that traversed the entire chamber, penetrating all of the foils. Such a track is shown in Fig. 15. While tracks such as this cannot be classed as measurable tracks because their range is not known, it seemed reasonable, in view of the identity of all other tracks observed, to assume these particles to be protons also. If this is true, these protons have energies in excess of 28 Mev.[7]

That such high energy protons are observed is interesting in that it indicates that, in some of the photo-disintegrations induced by high energy quanta, it is possible for a single particle to carry off a large fraction of the excitation energy in the nucleus, rather than this energy always being distributed among several particles.

Figure 16 is a plot showing the distribution in energy of the protons observed in this work. Further work is now being carried out to examine the tail of this distribution and determine the maximum energy of the photo-protons produced by 100-Mev x-rays.

[5] See following paper entitled "Cloud-chamber study of multiple scattering of protons in air."
[6] H. A. Bethe, Phys. Rev. **70**, 821 (1946).
[7] J. Smith, Phys. Rev. **71**, 32 (1947)

Cloud-Chamber Study of Multiple Scattering of Protons in Air*

E. A. LUEBKE, G. S. KLAIBER, AND G. C. BALDWIN
Research Laboratory, General Electric Company, Schenectady, New York
(Received February 10, 1947)

Experimental evidence is presented for occasional large apparent curvatures of short range proton tracks in a cloud chamber caused by multiple scattering in the gas. The observed curvatures agree qualitatively with the predictions of Bethe.

A METHOD in common use for determining the mass of a particle in cloud-chamber studies involves the simultaneous measurement of the range of the particle and the curvature of its path in a magnetic field. It has been pointed out by Bethe[1,2] that, in employing this method when dealing with particles having short ranges, multiple scattering of the particles by the atoms of the gas in the cloud chamber can cause large apparent curvatures of the tracks and thus introduce considerable error in the mass determinations. It is the purpose of this paper to present experimental evidence of the occurrence of these large apparent curvatures.

The experimental arrangement used in this work is the same as described in the preceding paper.[3] The cloud chamber was filled to atmospheric pressure with air and water and alcohol vapors, and a series of pictures was taken without a magnetic field being applied to the chamber. A total of 1180 photographs was taken and on these pictures there were 129 heavily ionizing tracks similar in character to the photo-protons described in the preceding paper. In accordance with the results of the previous work, it seemed legitimate to assume these 129 tracks were all due to protons.

Analysis of these 129 tracks gave the following results: 64 of the tracks were straight; 46 tracks showed visible curvatures approximating circular arcs, 21 of these curving in one direction and 25 curving in the opposite direction; 19 of the tracks exhibited an S-shaped curvature.

Of the 129 tracks, 18 were short ranged, stopping in the gas of the chamber before reaching the first foil. As predicted by Bethe,[2] these particles which stopped in the gas of the chamber showed instances of large apparent curvature. As the ranges of these particles stopping in the gas is known quite accurately, the observed curvatures, $1/r$, of the tracks can be compared with the root mean square curvatures, $1/r_{sc}$ due to scattering, as calculated according to Bethe.[2] Making this comparison, one finds 9 tracks having curvatures less than the calculated curvature, i.e., $1/r<1/r_{sc}$, and 6 tracks having curvatures greater than that calculated, i.e., $1/r>1/r_{sc}$. The remaining three tracks which terminated in the gas of the chamber exhibited S-curvature.

Tracks stopping in the first foil may have ranges between 15 cm and 44 cm, with corresponding values of r_{sc} in the upper part of the chamber of 170 cm and 490 cm; particles which stop at the center of this foil, having a range of 30 cm, should have an r_{sc} of 350 cm. Of the 21 tracks observed to end in this first foil, 6 showed S-curvature, 6 had $r\lesssim 350$ cm, and 9 had $r>350$ cm.

Several interesting cases of extreme curvature by multiple scattering are shown in the accompanying photographs. Figure 1 shows a direct (upper) and stereo (lower) view of a track with range 13 cm of air and an observed radius of curvature of 60 cm. This curvature exceeds the calculated root mean square curvature by a factor 2.5. Figure 2 is an enlarged view of the same track. A straight line is drawn near the track for comparison.

Another example of apparent curvature due to scattering is the track of shorter range presented in Fig. 3. It has a range of 10.5 cm air and an observed radius of curvature of 100 cm. This curvature is 1.2 times the calculated r.m.s. value.

* Presented in part at the September, 1946 American Physical Society Meeting in New York City.
[1] H. A. Bethe, Phys. Rev. 69, 689 (1946).
[2] H. A. Bethe, Phys. Rev. 70, 821 (1946).
[3] G. S. Klaiber, E. A. Luebke, and G. C. Baldwin, Phys. Rev. (preceding article).

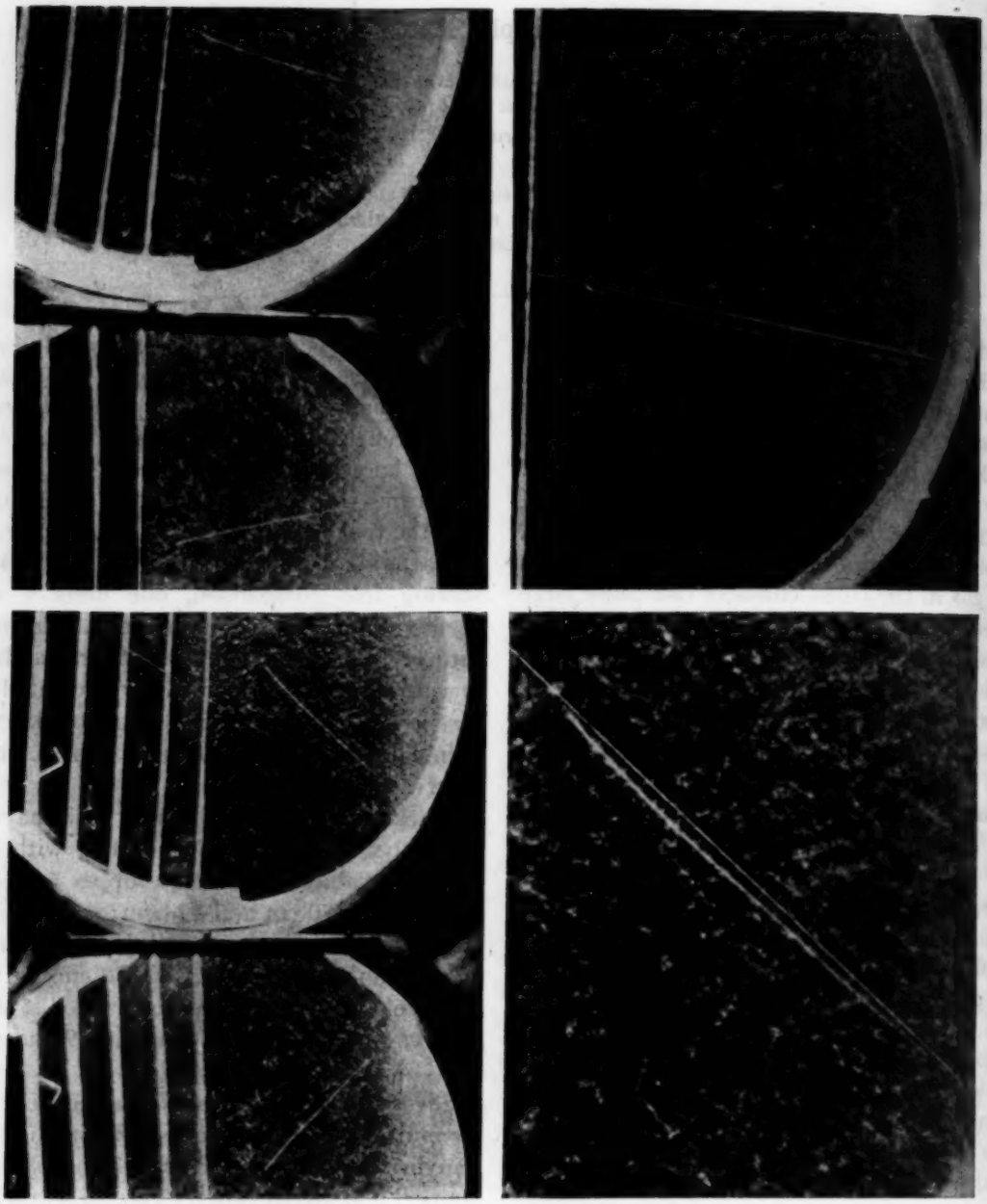


FIG. 1. Track of a proton of 13-cm range and 60-cm radius of curvature taken in zero magnetic field.
FIG. 2. Enlarged view of the track of Fig. 1 with straight line for comparison.
FIG. 3. Track of a proton of 10.5-cm range and 100-cm radius of curvature taken in zero magnetic field.
FIG. 4. Enlarged view of the shorter track of Fig. 3 with straight line for comparison.

The enlarged view of this track together with a straight line for comparison appears in Fig. 4.

The photographs of Figs. 5 and 6 present apparent curvature of the track of a particle that passes out of the illuminated portion of the cloud chamber. The range of this particle is greater than 14 cm air and the observed radius of curvature is 80 cm. This exceeds the calculated r.m.s. curvature by a factor greater than two.

Figure 7 is an example of the "S" curved track of a particle that stops in the first foil (range 12.5–42 cm air). Figure 8 shows this t rack with a

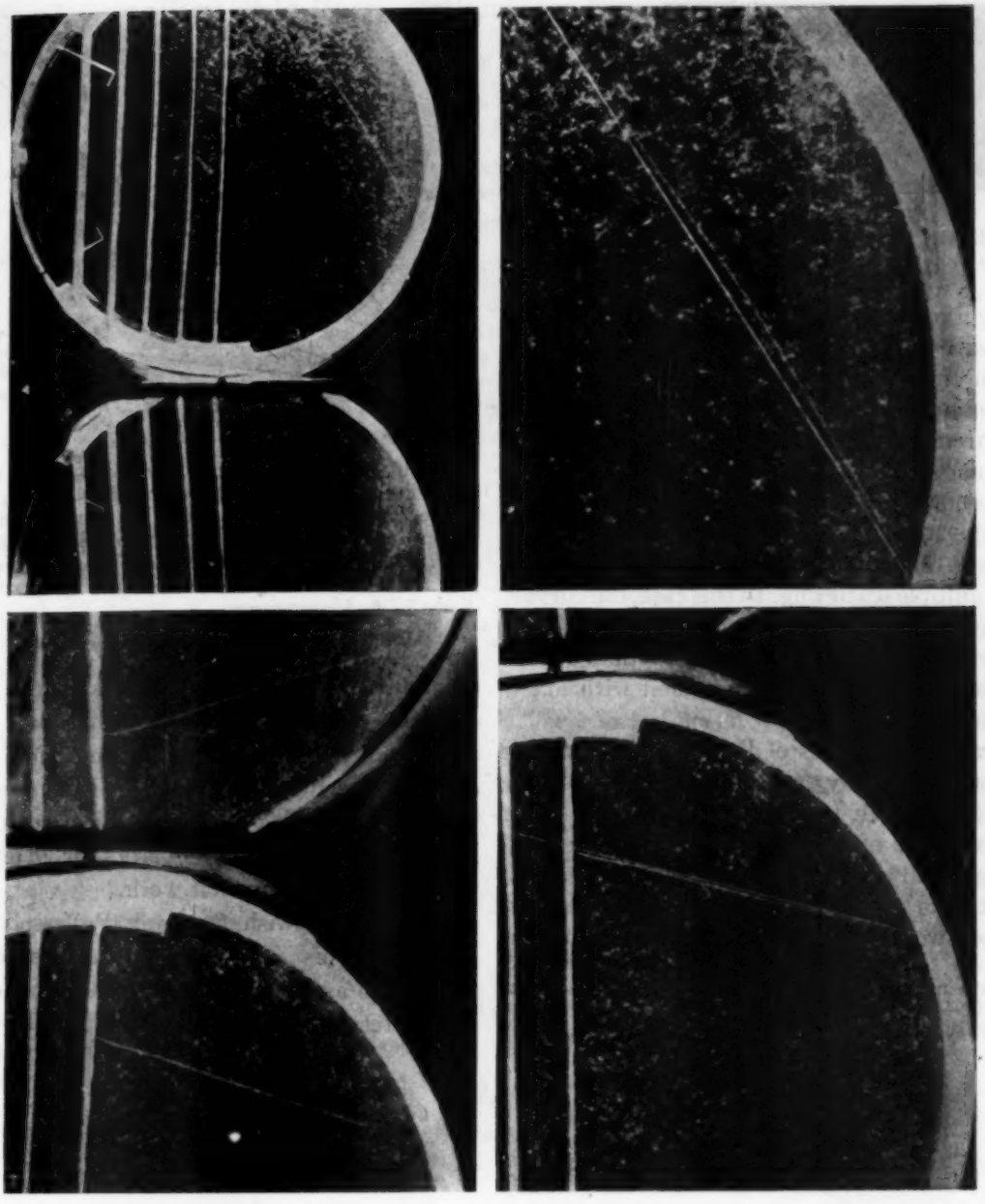


FIG. 5. Track of a proton with range in excess of 14 cm, having an 80-cm radius of curvature in zero magnetic field.
FIG. 6. Enlarged view of the track of Fig. 5 with straight line for comparison.
FIG. 7. S-curved track of a proton of range between 12.5- and 32-cm air taken in zero magnetic field. The middle 5 cm of the visible track has a radius of 120 cm.
FIG. 8. Enlarged view of the track of Fig. 7 with straight line for comparison.

straight line drawn along it for comparison. An extremely short range track with large apparent curvature is shown in Fig. 9. With range 3.5-cm air and radius of curvature 12 ± 3 cm, it has a curvature three times the calculated r.m.s. value.

In view of this evidence for apparent curvature of tracks due to scattering in the gas of the cloud chamber, it is of interest to consider the effect on the range-momentum measurements reported in the previous paper.[3] These effects are presented

FIG. 9. Heavily ionizing track of 3.5-cm range with 12±3-cm radius of curvature in zero magnetic field.

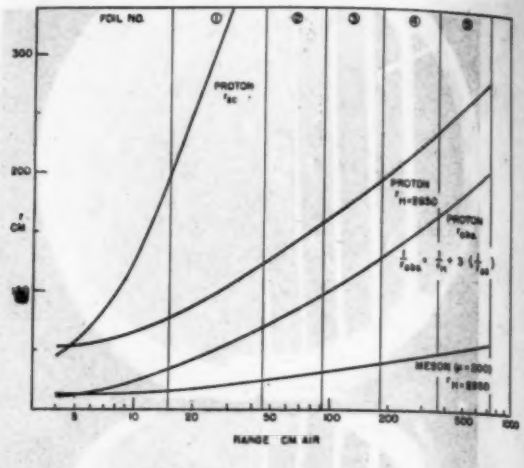


FIG. 10. Radii of curvature of proton tracks of various ranges in the experiment of the preceding paper: r_H, curved by the 2650 gauss magnetic field; r_{ss}, curved by multiple scattering in air; r_{obs}, curved by combination of magnetic field and extreme multiple scattering. These are compared with the radius of magnetic curvature of a meson track, shown in the lowest curve.

in the curves of Fig. 10. The ordinate is radius of curvature and the abscissa represents the range in cm air equivalent to the extent of the five aluminum foils present in the cloud chamber. The extreme case is considered where the observed curvature ($1/r_{obs}$) of a proton track is a combination of the magnetic curvature ($1/r_H$) due to the 2650 gauss field used in the experiment plus three times the calculated curvature ($1/r_{ss}$) due to multiple scattering. In this case, the curves indicate that protons can be distinguished from mesons of mass 200 for ranges greater than about 5-cm air equivalent in an experiment with magnetic field 2650 gauss.

The above examples of highly curved tracks are not sufficiently numerous for statistical analysis. However, they agree qualitatively with the predictions of Bethe and they indicate that the abnormally high curvature of several short range tracks which have appeared in recently published work[4,5] can be ascribed to the effects of multiple scattering in the gas of the cloud chamber. These curvatures, while not great enough to spoil the mass discrimination of the experiment with strong magnetic field described in the preceding paper, are sufficient to account satisfactorily for the few deviations from protonic mass encountered in that work.

The authors wish to thank Professor H. A. Bethe for his many helpful discussions on theoretical aspects, and Professor M. Schein for his interest in this work.

[4] M. Schein, A. J. Hartzler, and G. S. Klaiber, Phys. Rev. 70, 435 (1946).
[5] D. J. Hughes, Phys. Rev. 69, 371 (1946).

Gamma-Ray Yield Curve of Aluminum Bombarded with Protons

K. J. Broström and T. Huus
Institute for Theoretical Physics, University of Copenhagen, Denmark

AND

R. Tangen
Physical Institute, University of Oslo, Norway

(Received January 29, 1947)

The gamma-ray yield from Al^{27} bombarded with 0.5 to 1.4 Mev protons has been investigated with a sensitive G.M. counter. By introduction of a "cutter" device which automatically stopped the measurements when the voltage deviated more than a certain amount from the desired value, a resolving power of 4 kev was obtained. Under this condition a number of previously known peaks have been resolved into groups of narrow lines. In the investigated region the proton width has been estimated to be less than 1 kev, and a rough estimate of the efficiency of the gamma-ray counter gave a radiation width of the order of 1 ev.

INTRODUCTION

THE reaction $Al^{27}(p, \gamma)Si^{28}$ was first investigated by Herb, Kerst, and McKibben.[1] Using a thick target, they found a number of steps, which indicated the existence of resonance effects. Gentner[2] found two sharp resonances at 425 and 560 kev. Plain, Herb, Hudson, and Warren[3] made a detailed investigation of the excitation curve up to 2.6 Mev, using a thin target. They found a multitude of resonances. Hole, Holtsmark, and Tangen[4,5] have found four weak resonances between 220 and 500 kev.

Starting the work with the newly built Van de Graaff generator of the Institute for Theoretical Physics, Copenhagen, we intended to make a study of the excitation curve of this reaction, with the combined aim of matching the voltage scale with the results of the investigations mentioned above, and of finding new resonances with improved technique of detection.

EXPERIMENTAL

Generator

The pressure-insulated Van de Graaff generator was built with a grant from the Carlsberg Foundation to Professor Bohr. It is of the Lauritsen[6] type and will be described in greater detail at another occasion. The plans were made in 1940, and the collaboration of Dr. T. Lauritsen who at that time worked at this institute was of invaluable help in the design and construction.

The dimensions of the tank are 4.35×2.40 meters, and the maximum pressure is 7 atmospheres. The two belts are 50 cm wide. They have independent motors and spray volt plants. There are 52 corona rings, and 26 sections in the acceleration tube. The ion source is of the capillary

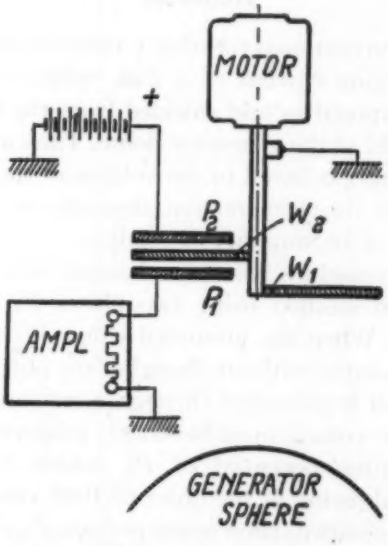


FIG. 1. Rotating compensation voltmeter.

[1] R. G. Herb, D. W. Kerst, and J. L. McKibben, Phys. Rev. 51, 691 (1937).
[2] W. Gentner, Zeits. f. Physik 107, 354 (1937).
[3] G. P. Plain, R. G. Herb, C. M. Hudson, and R. E. Warren, Phys. Rev. 57, 187 (1940).
[4] N. Hole, J. Holtsmark, and R. Tangen, Zeits. f. Physik 118, 48 (1941).
[5] R. Tangen, Thesis, Trondheim, 1947 (to be published).
[6] T. Lauritsen, C. C. Lauritsen, and W. A. Fowler, Phys. Rev. 59, 241 (1941).

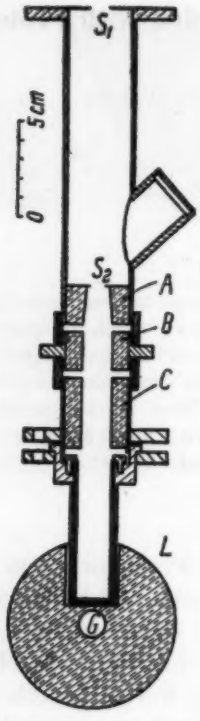


FIG. 2. Target arrangement.

type. The accelerated beam is analyzed magnetically.

A potential of approximately 1800 kv has been obtained with air at 6 atmospheres in the tank.

Voltmeter

The conventional rotating voltmeter measures the charging current of a disk, which is alternately exposed to and shielded from the electrostatic field of the generator dome. The current is directly proportional to the voltage of the generator, but its measurement depends on the reliability of an amplifier or rectifier.

This procedure has been changed into a compensation method using two plates P_1 and P_2 (Fig. 1). When the grounded wings W_1 and W_2 can be rotated without changing the potential of P_1, which is grounded through a resistance, the generator voltage must be strictly proportional to the positive potential of P_2, which is easily measured, being of the order of 1000 volts.

If the compensation is not perfect, P_1 will have a small a.c. voltage. This voltage is amplified and indicated on a small neon lamp of the type used as tuning indicators in wireless sets. The amplified signals from P_1 are also fed to the voltage stabilizer, which is of the McKibben type,[7] and to the "cutter," which blocks both gamma-ray counter and beam integrator when the voltage of the generator deviates more than a prescribed amount from its proper value.

The voltmeter is calibrated by means of the resonance in aluminum at 503 kev, which Tangen[5] has measured with high accuracy, both by comparison with the 440-kev resonance in lithium, and by an independent calibration of his voltage scale. The expected linearity is ascertained by measuring the 503-kev resonance with both protons and the molecular ions H_2^+ and H_3^+.

Beam and target arrangement

The ions were focused to a beam of 3 mm diameter and the current generally used was 1 to 2 μA protons. The current integrator is of the type used by Tangen.[5] It consists of a recorder which counts the pulses in a neon lamp, discharging a condenser constantly charged by the current.

The target arrangement is seen in Fig. 2. S_1 and S_2 are stops of tungsten foil. A, B, and C are lead cylinders, which serve to shield the gamma-ray counter against x-rays from the generator. The bore of these cylinders is greater than the diameter of the stop S_2, so that every ion coming through S_2 must strike the target. B is kept negative in relation to A and C, so that no secondary electrons from S_2 can pass to the target and *vice versa*.

The lowest part of the target tube can be removed for change of target. The target tube can be separated from the main vacuum system by a valve. The target was cooled by securing a good thermal contact between the target disk and the bottom of the lead box L.

The lead box contained a single G.M. counter placed as near to the target as possible. A small tube is preferable in the detection of weak radiation,[5] and a tube of 10-mm diameter and 40-mm length, was used. It was connected to a scale-of-32.

Preparation of targets

Targets were prepared by evaporation of aluminum in vacum on disks of copper or silver,

[7] D. B. Parkinson, R. G. Herb, E. J. Bernet, and J. L. McKibben, Phys. Rev. **53**, 642 (1938).

which had been given a fresh and clean surface by being turned carefully with a perfectly clean tool.

As a multitude of targets of different thicknesses had to be used, we have investigated the preparation of targets of known thicknesses in some detail. Because of the smallness of the resonance width, the resonances come out as approximately rectangular curves when the stopping power of the target is much greater than the inhomogeneity in proton energy, and the breadth of this rectangle immediately gives the stopping power of the target. We found the target thickness to be proportional to the amount of aluminum evaporated, so that targets of desired thicknesses could be produced with fairly good accuracy.

In order to estimate cross section values we also determined directly the amount of aluminum in a 25-kev target, and found a specific stopping power of about 250 kev per mg/cm² at 0.5 Mev-proton energy, in accordance with Parkinson, Herb, Bellamy, and Hudson.[8]

The aluminum had a specified purity of 99.998 percent. The targets will of course have an oxide cover, but all measurements of peak heights have been made on 10-kev targets, which we suppose to consist for the most part of metallic aluminum.

MEASUREMENTS

In determining the yield of the gamma-radiation over the energy interval examined, a great number of targets had to be used, since the targets were worn out after 2 to 3 hours of bombardment, even for a proton current of only 1 to 2 μA.

As a standard a target of about 1-kev stopping power at 500-kev proton energy was chosen. Targets of this type were used for all measurements on strong resonances (of intensities above 0.5 in the curve), whereas regions of no, or weak, radiation were searched with targets of about 4-kev stopping power. Finally a number of peaks throughout the spectrum were examined with targets thick enough (about 10 kev) to give saturation intensities. By these measurements the intensities obtained with the thinner targets, which could not be produced with exactly equal thicknesses, were brought to a common scale. The relative intensities of peaks are estimated to be reliable within 20 percent.

In most measurements the cutter was adjusted to allow voltage fluctuations of ± 1.5 kv. With this voltage definition a 1-kev target gives about $\frac{1}{3}$ of the saturation intensity.

The measurements on the 10-kev targets were also used for the establishment of the exact

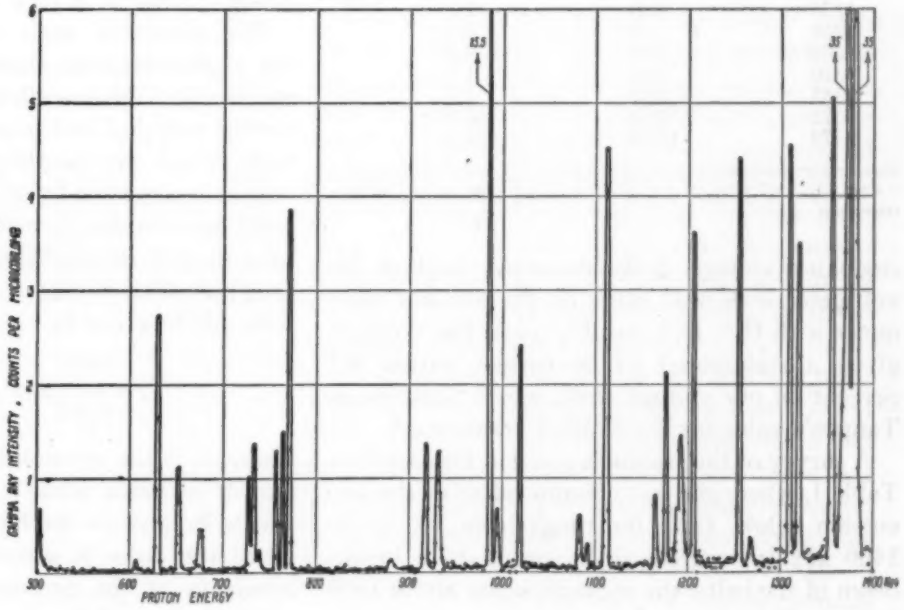


FIG. 3. Yield curve of aluminum bombarded with protons. All peaks have been plotted with the intensities found with the standard targets of about 1-kev stopping power ($\frac{1}{3}$ of saturation intensity), whereas the background has been plotted with the intensities found with 4-kev targets. The natural counting rate has not been subtracted.

[8] D. B. Parkinson, R. G. Herb, J. C. Bellamy, and C. M. Hudson, Phys. Rev. **52**, 75 (1937).

TABLE I. The first column gives the energy E in kev of the incident protons at resonance, while the second column contains the directly measured thick target yield N in counts per microcoulomb. The last column gives the quantity $\omega\gamma$ from formula (1), where ω generally is of the order $\frac{1}{2}$, and where γ, except for the weaker resonances, deviates only slightly from the radiation width, when the proton energy exceeds $E\sim 600$ kev.

E kev	N Counts per μC	$\omega\gamma$ ev
225*	0.005	0.4×10^{-3}
295*	0.015	1.3×10^{-3}
325*	0.080	6.0×10^{-3}
404*	0.30	25.0×10^{-3}
443*	0.050	5.0×10^{-3}
503	2.0	0.20
609	0.4	0.05
630	8.2	1.0
652	3.3	0.40
677	1.3	0.15
728	3.0	0.40
733	4.2	0.55
738	0.7	0.10
757	3.8	0.50
764	4.5	0.60
771	11.5	1.6
880	0.5	0.07
918	4.1	0.60
932	3.9	0.60
986	47.0	7.0
994	2.0	0.30
1018	7.2	1.1
1083	1.5	0.25
1091	0.8	0.12
1112	13.5	2.1
1165	2.4	0.40
1176	6.5	1.0
1192	complex	
1205	11.0	1.8
1255	13.0	2.1
1268	1.0	0.15
1309	14.0	2.3
1320	10.5	1.7
1355	15.0	2.5
1372	105.0	17.0
1379	105.0	17.0

* Measured by Tangen[5] and matched to our yield for the 503-kev resonance.

resonance voltage. A simultaneous check of the voltmeter scale was made on the 503-kev resonance with H_1^+, H_2^+, and H_3^+ ions. The voltages given are supposed to be correct within 0.2 percent on our voltage scale, which depends on Tangen's value for the 503-kev resonance.[5]

A survey of the results is given in Fig. 3 and in Table I, where the last column refers to the discussion below. Only the range from 500 up to 1400 kev is covered, since, owing to a breakdown of the belts, the measurements above 1400 kev are so far incomplete.

The curve shows 31 prominent peaks, 16 of which are found in the curve of Herb *et al.*[3] The resonance at 862 kev, which Herb ascribes to fluorine, has not been found, even though only weak background radiation is present in that region. As no considerably stronger fluorine lines exist in the examined voltage region, the curve can therefore not be influenced by fluorine contamination. The complex peak at 1192 kev seems to consist of three lines at 1189, 1192, and 1195 kev, but nothing definitely can be said.

The gradual rise of the background with increasing voltage is mainly due to radiation from the silver disk. Copper gives considerable radiation above 1000 kev and has therefore not been used as target support at higher voltages.

DISCUSSION

According to the general scheme for nuclear reactions, the observed resonances correspond to stationary states of the compound nucleus Si^{28} formed and excited to about 12 Mev by the impact of a proton on a nucleus of Al^{27}, which is the only isotope present in natural aluminum. In the present case, where neutron escape is excluded for energetic reasons, and where the emission of an α-particle is very improbable because of the large barrier, the final result of the collision will depend on a competition between the re-emission of a proton and radiation processes leading to a final capture of the proton.

The yields of such resonances depend on the probabilities for these competing processes, specified by the so-called widths Γ_p and Γ_γ for proton emission and gamma-radiation, respectively, which are simply the probabilities measured in energy units by multiplication by Planck's constant over 2π. According to the well-known formula of Breit and Wigner, the cross section for (p,γ) processes integrated over one of the narrow lines will be given by

$$\Sigma=\frac{h^2}{4ME}\omega\gamma;\quad \gamma=\frac{\Gamma_p\Gamma_\gamma}{\Gamma_p+\Gamma_\gamma}, \tag{1}$$

where E is the resonance energy of the proton and M its mass, while ω is a spin weight factor which, because of the high spin 5/2 of Al^{27}, will be of the order $\frac{1}{2}$, although, for large angular momenta of the protons, it may deviate appreciably from this value.

The thick target gamma-ray yield of a single

resonance, i.e., the fraction Y of the incident protons finally captured through the radiation processes, will therefore be given by the expression

$$Y=\frac{\sum}{sMA}=\frac{s_0\omega\gamma}{sAE}, \qquad (2)$$

where A is the mass number and s the specific stopping power of the target material, while s_0 is an abbreviation for $h^2/4M^2$ and equal to 2.5 kev per mg/cm^2.

For the specific stopping power of aluminum we have used the values given by Herb *et al.*,[8] which for 500 kev agree with our measurements.

By comparing the yield from fluorine with the absolute yield given by Lauritsen *et al.*,[9] and by van Allen and Smith,[10] we find the efficiency of the counting arrangement to be 0.12 percent for 6 Mev quanta. This corresponds approximately to 0.2 percent for the aluminum radiation. The estimate of the efficiency is rather rough and constitutes the main source of uncertainty in the values, given in Table I, of the quantity $\omega\gamma$ calculated by means of Eq. (2).

From Eq. (1) it is seen that γ will be of the same order of magnitude as the smallest of the two partial widths Γ_p and Γ_r. In the low voltage region where, due to the influence of the barrier, Γ_p is very small compared with Γ_r but increases rapidly with the proton energy, we should therefore, according to Eq. (2), expect the resonances to rise gradually in strength with increasing energy E.

In the high voltage region we may, of course, still find small values of Γ_p corresponding to the impact of protons with high angular momenta, but the value of Γ_p for zero angular momentum should be large compared with Γ_r. On entering this region the general increase in the resonance strength should therefore be expected to stop, since the radiation width is normally assumed to be approximately constant.

Such a dependency on the proton energy is also indicated by the general trend of the yield curve (Fig. 3). The transition region between "low" and "high" energies appears to be situated in the vicinity of 600 kev. Below this energy the values given in the third column of Table I should thus represent $\omega\Gamma_p$, while for higher energy they should represent $\omega\Gamma_r$, except in cases where we have to do with comparatively large angular momenta of the incident protons.

Since ω will on the average be about $\frac{1}{2}$, one finds values of Γ_r of the order of magnitude of 1 ev, in accordance with general theoretical estimates. For the three very strong levels at 986, 1372 and 1379 kev it seems necessary, however, to assume excessively large values of Γ_r, which might perhaps be caused by some specific character of the excitation states favorable for radiative transitions.

As regards the order of magnitude of the proton width, we may assume that for zero angular momentum it will rise rapidly from about 1 ev for $E\sim600$ kev, to values of about 1 kev at 1400 kev, as estimated from the penetrability of the barrier. Thus, for none of the resonances is the natural width, which is the sum of the partial widths, revealed by the measurements, but still the slope of the foot of the strong double-line at 1372 and 1379 kev may give an indication of the actual width. In fact, by comparing the shape of the curve in this region with the foot of an ordinary dispersion line with the given area, one arrives at a total width of just about 1 kev.

It need hardly be added that the "foot" might originate from one or more weak lines which have escaped detection by the present resolution, but there is reason to believe that the cutter arrangement is sufficiently reliable to prevent any appreciable influence from the central part of the line at the distance in question, and this interpretation is also confirmed by the sharpness of the strong line at 986 kev for which, with the estimated much smaller proton width, no measurable foot should be expected.

In the region from 600 to 1400 kev the mean distance between resonances is seen from Table I to be about 30 kev. The lines are within this region fairly evenly distributed, but there is a remarkable grouping and perhaps some regularity in the spacing of the levels of each group (e.g., the three lines at 757, 764 and 771 kev), although, it cannot be excluded that some of the coincidences are of a more accidental character. Still, especially in the case of the two very strong lines at 1372

[9] J. F. Streib, W. A. Fowler, and C. C. Lauritsen, Phys. Rev. 59, 253 (1941).
[10] J. A. van Allen and N. M. Smith, Jr., Phys. Rev. 59, 501 (1941).

and 1379 kev, it is difficult to escape the impression of the presence of a typical fine structure.

In a forthcoming publication from this institute it is planned to give a more detailed discussion of the characteristics of the spectrum.

The authors wish to express their thanks to Professor Niels Bohr for offering them the facilities to carry out this work, and for his valuable help with the discussion of the results. One of us (R. T.) is also indebted to "The Foundation for Danish-Norwegian Co-operation" and the Norwegian "Foundation of 1919 for Scientific Research" for grants enabling him to take part in the work in Copenhagen.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

Interference Phenomena of Slow Neutrons

E. FERMI AND L. MARSHALL
Argonne National Laboratory and University of Chicago, Chicago, Illinois

(Received February 7, 1947)

Various experiments involving interference of slow neutrons have been performed in order to determine the phase of the scattered neutron wave with respect to the primary neutron wave. Theoretically this phase change is very close to either 0° or 180°. The experiments show that with few exceptions the latter is the case. The evidence is based on the following types of measurements: (a) measurement of the intensities of Bragg reflection of various orders of many crystals, and comparison with the theoretical values of the form factor; (b) total scattering cross section of gas molecules for wave-lengths long compared with the molecular dimensions; and (c) determination of the limiting angles for total reflection of neutrons on various mirrors. The elements Ba, Be, C, Ca, Cu, F, Fe, Mg, N, Ni, O, Pb, S, and Zn were found to scatter neutrons with 180° phase difference; Li and probably Mn scatter with zero phase difference. The five elements I, Br, Cl, K, and Na behave alike and the phase with which they scatter is tentatively identified as 180°. Coherent scattering cross sections have been determined for several elements.

INTRODUCTION

THE scattering processes of slow neutrons are greatly complicated by interference phenomena due to the fact that the de Broglie wave-length is comparable with interatomic distances. The general pattern of interference phenomena of slow neutrons is similar to that of x-rays, since both the wave-length and the scattering cross section of x-rays are comparable to those of neutrons. On the other hand there are considerable differences due to several factors. Among them is the fact that the scattering of x-rays varies regularly with atomic number while that of neutrons is a rather erratic property. Furthermore, in the case of neutrons the phase difference between scattered and incident wave may be either 0° or 180° as will be discussed in Section 1. For x-rays instead, it is always 180° because x-ray energies are larger than most electronic resonance energies. Also, the absorption properties of neutrons differ markedly from those of x-rays.

The main purpose of this work was the investigation of various interference phenomena in order to determine the phase change of the scattered neutron wave for a large number of elements. Section 1 contains a summary of the theoretical background of this work. Section 2 describes the measurements of the intensities of Bragg reflections of various orders and their interpretation. Section 3 is a discussion of some experiments on filtered neutrons. In Section 4, experiments on scattering of neutrons by gas molecules are presented. Section 5 describes measurements of the limiting angle for total reflection of neutrons. The general conclusions are discussed in Section 6.

1. Theoretical Considerations

When a slow neutron is scattered by a nucleus, its de Broglie wave at some distance from the scattering nucleus may be written as the sum of a term $\exp(ikx)$ representing the primary wave and a term $-a[\exp(ikr)]/r$ representing the scattered wave. This is true for the case of slow neutrons because with very good approximation the scattering is spherically symmetrical. The coefficient of the latter term has been written with the minus sign for reasons of convenience that will be apparent later. If the constant a is positive, there is a phase change of 180° between scattered and incident wave, and if a is negative, the phase change is 0°. One proves in an elementary way that the scattering cross section is related to the constant a by the equation

$$\sigma = 4\pi |a|^2 \tag{1}$$

and also that the constant a is with very good approximation for slow neutrons a real number. Its imaginary part is very small and can usually be neglected except in case of extremely high absorption. The quantity a which has the dimensions of a length and the order of magnitude of 10^{-12} cm shall be referred to as the "scattering length." The main purpose of this paper is the experimental determination of the scattering length and in particular of its sign.

For elements consisting of one isotope only and without nuclear spin, the magnitude of the scattering length can be immediately obtained from (1), so that only its sign needs to be determined. Even in this simple case a small correction must be applied depending on whether the atom is free or bound. Because of the change in the reduced mass, the cross section of a free atom differs from that of the same atom bound in a crystal, by a factor $[(A+1)/A]^2$, where A is the atomic weight.[1]

Since the scattering length is proportional to the square root of the cross section, the correction factor will be $(A+1)/A$. It follows that

$$a = [(A+1)/A]a_f \tag{2}$$

where a indicates the scattering length of the atom bound in a crystal and a_f the scattering length for the free atom. For elements that are mixtures of several isotopes with probabilities of occurrence $p_1, p_2, \cdots, p_n$, and scattering lengths $a_1, a_2, \cdots, a_n$ the scattering length is the average

$$a = p_1a_1 + p_2a_2 + \cdots + p_na_n. \tag{3}$$

This is the magnitude that determines the interference properties of the element. The scattering cross section is no longer given by (1) but by

$$\sigma = 4\pi\{p_1a_1^2 + p_2a_2^2 + \cdots + p_na_n^2\}. \tag{4}$$

Actually in this case the relationship between a and σ can be expressed by the inequality

$$|a| \leq (\sigma/4\pi)^{\frac{1}{2}}. \tag{5}$$

The knowledge of the scattering cross section is insufficient, therefore, to determine even the magnitude of the scattering length.

A similar situation obtains if the nucleus has a spin I different from zero. Here the spin vector of the scattering nucleus is oriented either parallel (total spin $I+\frac{1}{2}$), or antiparallel (total spin $I-\frac{1}{2}$) to the spin vector of the neutron. Consequently the scattering length may take either of two values, $a_{I+\frac{1}{2}}$ or $a_{I-\frac{1}{2}}$. The effective scattering length is the average of these two scattering lengths, each being weighted by the probability of occurrence of the corresponding spin orientation, namely

$$a = \frac{I}{2I+1}a_{I-\frac{1}{2}} + \frac{I+1}{2I+1}a_{I+\frac{1}{2}}. \tag{6}$$

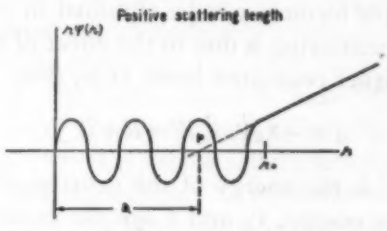


FIG. 1A. Positive scattering length.

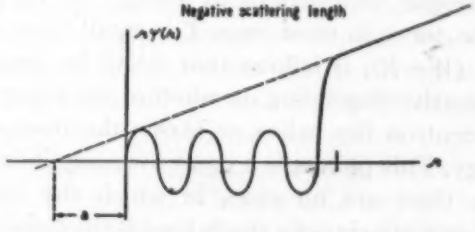


FIG. 1B. Negative scattering length.

[1] For example see H. A. Bethe, Rev. Mod. Phys. 9, 71 (1937).

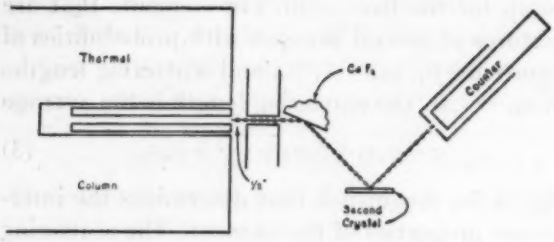


FIG. 2. Apparatus for measuring intensity of Bragg orders.

The scattering length will obey (5), the equal sign corresponding to the case that $a_{I+\frac{1}{2}}$ is equal to $a_{I-\frac{1}{2}}$ both in magnitude and sign.

A simple geometrical interpretation of the scattering length can be obtained if the interaction between the neutron and the nucleus is represented by a potential well. Let the eigenfunction of the neutron for an s state of energy zero be $\psi(r)$. In Fig. 1, $r\psi(r)$ is plotted *versus* r. The curve becomes a straight line for r greater than the radius r_0 of the potential well. This straight line can be extended to intersect the r axis at the point P. One can prove that the scattering length a is the abscissa of the point P. Figure 1A corresponds to a case in which a is positive, and Fig. 1B to a case in which a is negative. By inspecting the two figures it is apparent that the probability is much higher for a positive than for a negative, especially for heavy elements. This remark on the relative probability for a to be positive or negative is due to E. Teller and V. Weisskopf.

A simple formula can be obtained in the case that the scattering is due to the effect of a single Breit-Wigner resonance level. It is, then

$$a=-\lambda_R\Gamma_n/[W-R+i\Gamma] \qquad (7)$$

where W is the energy of the neutron, R is the resonance energy, Γ_n and Γ are the neutron and total half width at half maximum, and λ_R is the de Broglie wave-length at energy R, divided by 2π. Since in most cases Γ is small compared with $(W-R)$, it follows that a will be positive or negative depending on whether the energy of the neutron lies below or above the resonance energy. This picture is a great oversimplification since there are no cases in which the Breit-Wigner scattering of a single level is the dominant phenomenon, and complicated situations arise due to interference of Breit-Wigner and potential scattering.

2. Intensities of Various Orders of Bragg Reflection

One can compare the changes in phase in the neutron scattering of different elements by measuring the intensity of the Bragg scattering for various orders and various crystalline planes of crystals containing at least two elements.

The simplest case is that of a crystal in which the planes are equidistant and consist alternately of two kinds of atoms, as for example, the 1, 1, 1 planes of NaCl. These planes are equidistant and consist alternately of sodium and of chlorine. In the first order Bragg reflection, the optical path for reflection from sodium planes differs from the path for reflection from chlorine planes by $\lambda/2$. Consequently if sodium and chlorine nuclei cause the same change in phase of the scattered neutron wave, their contributions will subtract and the order will have low intensity. If they scatter with opposite phase change their contributions will add and the order will have high intensity.

The situation for the second order is reversed. Here the difference in optical path for reflection from the two kinds of planes is λ. Consequently if sodium and chlorine scatter with the same change in phase the order will have high intensity, and if with opposite change in phase, low intensity. Similarly it follows that if the two kinds of nuclei scatter with the same change in phase the odd orders will be weak and the even orders will be strong. Conversely, if they scatter with opposite change in phase, the odd orders will be strong and the even orders weak.

Analogously to x-ray scattering, there is superimposed upon this effect a continuous decrease in intensity from order to order due to geometrical factors, to thermal agitation, and to imperfections of the crystal.

For more complicated cases the intensity of the various orders is determined as for x-rays by the form factor

$$F=|\sum_j a_j \exp(2\pi i n\delta_j/d)| \qquad (8)$$

where a_j is the scattering length, d is the spacing of the lattice planes, n is the order of the Bragg

reflection, and δ_j is the perpendicular distance from the jth atom to the plane of reflection. Naturally the form factor is strongly dependent on the relative signs of the scattering lengths of the various atoms of the crystal. An analysis of the intensities for various orders and various planes of the crystal will often allow the determination of the relative signs of the scattering lengths of the elements involved. In several cases it is possible to determine also the ratios of the values of the scattering lengths.

A comparison of intensities for Bragg reflections of various orders of many crystals have been made using monochromatic neutrons. The experimental arrangement is schematically represented in Fig. 2. A beam of non-monochromatic thermal neutrons emerges from a $4'' \times 4''$ hole containing a long collimator $\frac{1}{2}''$ wide by $1\frac{1}{4}''$ high in the thermal column of the Argonne heavy water pile. This beam falls on the 1, 0, 0 plane of a large CaF_2 crystal at a glancing angle of about 16°. Neutrons of energies satisfying the Bragg condition ($\lambda \sim 1.50$ angstroms for the first order) are reflected to form a beam that is used for investigating various crystal specimens. The beam contains a main component of neutrons of wave-length 1.50A plus a small fraction of neutrons of energy four times as large due to the second order reflection on the CaF_2 crystal. The second order component is fairly unimportant because there is only a small number of neutrons of this relatively high energy in the Maxwell distribution. We can, therefore, consider the beam reflected on the calcite to be approximately monochromatic. This monochromatic beam of neutrons falls on some plane of the second crystal, which is mounted on a rotating table. The neutrons undergo a second Bragg reflection, and are finally detected by an enriched BF_3 proportional counter. The counter is supported by an arm which rotates about the axis of the crystal table. Additional Cd slits not shown are used for more precise collimation of the neutron beam.

A typical measurement was carried out as follows. Both counter and second crystal were set approximately at angles corresponding to a reflection of a given order. By rocking each separately the setting of maximum intensity was found. Also the inclination of the crystal was adjusted for maximum intensity. From the intensity measured in these conditions was subtracted a background, measured by turning the crystal a small angle off the Bragg position. In the majority of cases the background was a small fraction of the total count.

The results of these measurements are summarized in Table I. Columns 1 and 2 indicate the crystal and the plane. Column 3 is the form factor. Here the chemical symbol of the element has been used to represent its scattering length. The fourth column gives the absolute value of the form factor, using the values of the scattering length given in Table VII. The fifth column gives the measured intensities of the various orders in counts per minute. In the sixth column is the ratio of column 5 to column 4. If the values chosen for the scattering lengths are approximately correct, one expects that the values in column 6, corresponding to any given plane, should show a regular decrease in intensity with increasing order, since dividing the intensity by the form factor should correct it for the irregular change of intensity from order to order. Therefore only the regular decrease should remain.

There may be some doubt whether it is more appropriate to divide the intensity by the form factor or by its square. Theoretically one would expect that, for ideally perfect crystals, the form factor should be used, and for ideally imperfect crystals, its square. Actually, we have found that one obtains a much better fit by using the form factor.

For the simplest crystals containing two elements one can see immediately by inspecting the intensity data of column 5 whether the scattering lengths of the two elements have equal or opposite sign. For instance CaF_2 (100), NaCl (111), and PbS (111) show clearly superimposed on the general decrease of intensity with increasing order, an alternation of intensity with strong even orders and weak odd orders. Since, in these crystals, the planes consist alternately of the two kinds of atoms, we conclude that the scattering lengths of each pair have the same sign. The opposite is the case for LiF (111) where the even orders have intensity so low that we could not measure them, and the odd orders have a normal intensity. Therefore Li and F have scattering lengths of opposite sign. In some of the more complicated cases the analysis is also

TABLE I. Intensity of Bragg orders.

Crystal	Plane	Order		Form factor		Intensity	Ratio
CaF_2	(100)	1		Ca−2 F	.41	16300	39800
		2		Ca+2 F	1.99	20300	10200
		3		Ca−2 F	.41	1287	3140
NaCl	(111)	1		Na−Cl	.57	2376	3990
		2		Na+Cl	1.69	2750	1630
PbS	(111)	1		Pb−S	.20	7280	36400
		2		Pb+S	.76	10700	14100
		3		Pb−S	.20	808	4040
		4		Pb+S	.76	750	986
PbS	(100)	1		Pb+S	.76	19650	25800
		2		Pb+S	.76	11420	15000
		3		Pb+S	.76	2249	2960
FeS_2	(100)	1		Fe+ .62 S	.99	6893	6960
		2		Fe−1.6 S	.37	1354	3660
		3		Fe−1.6 S	.37	726	1960
MnS_2	(111)	1		Mn−1.06 S	.74	7930	10700
		2		Mn+ .06 S	.42	2560	6100
		3	M	Mn+ .06 S	.42	670	1600
Fe_3O_4 (magnetite)	(111)	1		2.34 Fe+ 1.20 O	2.65	21300	8040
		2		16 Fe−31.9 O	6.34	19970	3150
		3		13.7 Fe− 3.55 O	9.06	12570	1390
		4		8 Fe+31.50 O	25.78	16700	647
		5		13.7 Fe+ 5.94 O	14.85	5120	344
$CaCO_3$ (calcite)	(111)	1		2 C+6 O−2 Ca	3.42	9500	2780
		2		2 C+6 O+2 Ca	6.58	8520	1290
		3		2 C+6 O−2 Ca	3.42	1840	538
$CaCO_3$ cleavage	(211)	1		2 Ca+2 C+1.76 O	3.99	10900	2730
		2		2 Ca+2 C−2.0 O	1.70	3360	1980
		3		2 Ca+2 C+2.7 O	4.57	2400	528

quite straight forward as can be seen by comparison of the observed intensities with the formulae for the form factors in column 3. In a few cases the interpretation is not unique.

The following conclusions as to the signs of the scattering lengths can be drawn. The components of the following pairs have scattering lengths of the same sign: Na, Cl; Pb, S; Ca, F; Fe, S; Fe, O; K, Br; Mg, O; K, Cl; K, I. The components of the pair, Li, F have scattering lengths of opposite sign. The measurements of MnS_2 do not allow a unique interpretation although the probable conclusion is that Mn and S also have scattering lengths of opposite sign.

Using the evidence to be presented in Sections 4 and 5 that carbon and oxygen have scattering lengths with the same sign, one can conclude from the measurements on calcite that Ca and O also have scattering lengths of the same sign.

From the measurements on FeS_2 and Fe_3O_4, one can conclude that S and O have the same sign; and from data on $BaSO_4$, one is led to assign also the same sign to Ba. The measurements on $NaNO_3$ indicate that N and O have the same sign. (See confirmatory evidence in Section 4.) The measurements of the (111) plane of $NaNO_3$ indicate although not quite conclusively that Na has the same sign as the group NO_3, and therefore as N and O.

From these data it follows that C, O, Fe, Mg, Ba, Ca, S, F, Pb, and N all have the same sign. Na, K, Cl, Br, I all have like sign. The partial evidence just mentioned, according to which Na has the same sign as O, indicates that all this latter group should be included with the former. That this conclusion is correct is strongly supported by evidence to be presented in Section 6. Li and perhaps Mn have scattering lengths of the other sign. They represent together with hydrogen (see Section 6) the only exceptions

TABLE I.—*Continued.*

Crystal	Plane	Order	Form factor		Intensity	Ratio
$BaSO_4$	(100)	1	4 Ba+4 S+ .42 O	4.54	3606	794
		2	4 Ba+4 S+14.6 O	13.19	4837	366
		3	4 Ba+4 S+ 3.1 O	6.17	756	123
$BaSO_4$	(011)	1	−3.4 Ba+3.8 S+4.7 O	1.24	1018	821
		2	1.75 Ba+3.1 S−9.6 O	3.61	1632	452
		3	.42 Ba+2.0 S−7.9 O	3.93	651	166
		4	−2.46 Ba+ .75 S−2.2 O	3.07	36	12
$BaSO_4$	(101)	1	.87 Ba−1.25 S− 1.47 O	.56	737	1320
		2	1.13 Ba−2.02 S− 4.56 O	2.46	1946	791
		3	3.77 Ba+ .98 S+ 2.56 O	4.81	1477	307
		4	.27 Ba− .10 S− 2.12 O	1.47	—	—
		5	1.04 Ba+2.12 S−10.4 O	4.93	587	119
MgO	(100)	1	Mg+O	1.17	10352	8800
		2	Mg+O	1.17	6258	5350
MgO	(111)	1	Mg−O	.05	764	15300
		2	Mg+O	1.17	14175	12100
		3	Mg−O	.05	132	2640
LiF	(111)	1	Li−F	1.19	10080	8470
		2	Li+F	.01	~0	—
		3	Li−F	1.19	300	252
KBr	(111)	1	Br−K	.21	1545	7360
		2	Br+K	.91	2853	3140
KCl	(111)	1	Cl−K	.78	2346	3010
		2	Cl+K	1.48	5620	3800
		3	Cl−K	.78	334	428
		4	Cl+K	1.48	160	108
$NaNO_3$	(211)	1	Na+N+ .88 O	1.96	29000	14800
		2	Na+N− .98 O	.83	4171	5000
		3	Na+N+1.34 O	2.25	4148	1800
$NaNO_3$	(111)	1	NO_3−Na	2.14	29400	13700
		2	NO_3+Na	3.26	16790	5150
		3	NO_3−Na	2.14	2243	1050
KI	(111)	1	K−I	.01	85	—
		2	K+I	.71	348	500
		3	K−I	.01	24	—

found so far to the behavior of the majority of elements.

This type of data allows only a comparison of the signs of scattering lengths of different elements but not their absolute determination. The absolute determination of sign of scattering length will be discussed in Section 5. The over-all conclusions from Section 2 will be presented in Table VII. There, an attempt has been made to give the actual value of the scattering lengths for several elements. In calculating these values the attempt was made to obtain the best possible agreement of the observed intensities given in Table I, with the form factor after proper account was taken of the natural decrease of intensity with increasing order.

In x-ray analysis of crystals the Debye-Scherrer method of powder photography is extensively used. For neutrons, however, the intensity of the Debye-Scherrer maxima is rather small, and in most cases the method will be impractical. For example, in the case of microcrystalline graphite, we were able to detect only the maximum corresponding to the first reflection on the 001 plane.

A second possibility to study microcrystalline substances is to measure the total cross section of the substance for neutrons of wave-length

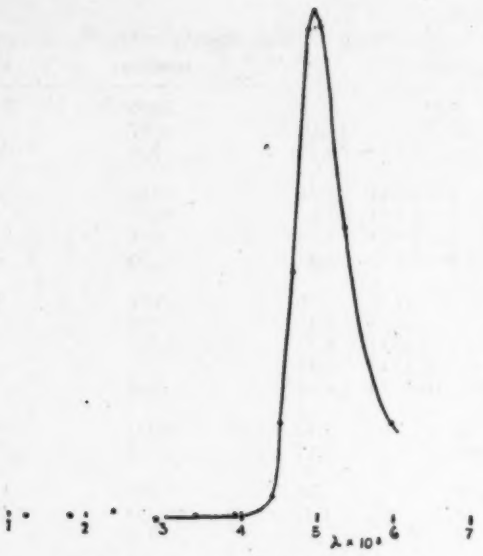


FIG. 3. Spectrum filtered through BeO.

close to $2d$ where d is the longest lattice spacing. An analysis of this case for substances containing two elements has been made by Fermi and Sachs, and indicates that the total cross section in this region depends strongly on whether the scattering lengths of the two elements have equal or opposite sign. The practicability of this method is somewhat limited by the fact that the theory applies only when the microcrystalline grain of the substance is exceedingly fine.

3. Spectrum of Filtered Neutrons

In the study of interference in gas molecules, somewhat simpler results are obtained with neutrons of energy far below thermal energies and which have a wave-length long compared with interatomic distances. Such neutrons are present in very small percentage in the Maxwell distribution at room temperature. Consequently, one cannot isolate such low energy neutrons by Bragg reflection because in the reflected beam there is a contamination by high orders which is many times more intensive than the first order. A simple way to avoid this difficulty, without use of extremely low temperature moderators, is to remove the high energy part of the Maxwell distribution by passing the beam of neutrons through a filter of a microcrystalline material of low absorption,[2] such as graphite or BeO. This filtering action of microcrystalline substances is due to the fact that neutrons of wave-length longer than $2d$, where d is the longest lattice spacing of the crystal, cannot satisfy the Bragg reflection condition for any of the many crystallites that they meet while crossing the filter. Consequently, they are scattered out of the beam only to a minimal extent. The production of very low energy neutrons by the filtering process has been previously reported[2] for the case of graphite, where the limiting wave-length is 6.7A. In the present experiments BeO filters have been used, for which the limiting wave-length is 4.4A. In order to get experimental proof of the correctness of the interpretation of the filtering process, we have investigated with the crystal spectrometer the spectrum of the neutrons filtered by a prism 40 cm long and containing 100 g/cm². The side dimensions of the prism were about 10×10 cm. The analysis was carried out with a crystal of celestite in the 001 plane. The results are plotted in Fig. 3, where the wave-lengths are plotted on the abscissae, and the observed intensities are the ordinates. The figure shows clearly that the spectrum of the filtered neutrons has negligible intensity for short wave-length, rises abruptly to a maximum value at wave-length of approximately 4.5A, and after that, decreases gradually with increasing wave-length.

The decrease of intensity on the long wave-length side is somewhat more rapid than corresponds to the Maxwell distribution. This probably is due to the fact that the neutrons emerging from the graphite column are not fully slowed down to the lowest energies of the Maxwell distribution by the collisions in graphite.

The fact that microcrystalline BeO and Be metal are good filters for neutrons is an indication that the scattering lengths of Be for the two spin orientations of the neutron do not differ very considerably. Indeed if this were the case one would expect a strong incoherent scattering that should not vanish even for wave-lengths longer than $2d$. For filtered neutrons we found a residual cross section for BeO of 0.7×10^{-24} cm². If all this residual cross section were due to the difference in the scattering lengths a_1 and a_2 for the two spin orientations, it could be calculated that a_1 and a_2 certainly have the same sign and that their ratio is about 2. Probably the actual

[2] Anderson, Fermi, and Marshall, Phys. Rev. 70, 815 (1946).

values are appreciably closer, because part of the observed residual cross section is certainly due to crystal imperfections and to thermal agitation.

4. Scattering of Neutrons by Gas Molecules

Some conclusions on the scattering of neutrons can be drawn from the study of the total cross section of gas molecules. For neutrons of wave-length short compared with interatomic distances the molecular scattering cross section is the sum of the individual scattering cross sections of the atoms in the molecule. When neutrons of longer wave-length are used, a number of complications arise, partly due to interference effects, and partly to the fact that the scattering now is due to atoms that no longer can be considered free.

The theoretical calculation of the cross section has been carried out by Teller and Schwinger[3] for the case of the H_2 molecule for low rotational states. For heavier molecules, such as N_2 or O_2, even when low temperatures and slow neutrons are used, a fairly large number of rotational states is always excited so that a detailed calculation of the contributions of all rotational states becomes impractical. Simple results can be calculated by neglecting the neutron mass in comparison with the mass of the atoms in the molecule. In this approximation one obtains results identical with those of the classical theory of interference. The approximation is quite good in the case of x-rays, where the particle scattered is a photon of very small effective mass, and has been used currently by Debye and his co-workers. For scattering of neutrons from molecules such as N_2, O_2, etc., the approximation is not nearly as good. The corrections, however, have not been calculated, and the results will be compared with those of the approximate theory except in the case of H_2 where the theory of Teller and Schwinger has been used.

The scattering cross sections of several molecules have been determined using very low energy neutrons obtained by filtration through a BeO filter. The spectrum of these neutrons is given in Fig. 3. Although these neutrons are not monochromatic they belong to a fairly narrow band of average wave-length 5.1A. For the measurements at room temperature the gas under investigation was contained in a long aluminum tube of 2.5" i.d. and 365 cm long. Pressures up to 2 atmospheres were used. On H_2 we performed some measurements at liquid air temperature and in this case a tube 24" long was used. A copper coil was soldered around this tube and liquid air was circulated in it. The entire assembly was protected with rock wool insulation. The conditions were such that the ortho-para ratio was practically the same as at room temperature.

Only in the cases of CF_4 and H_2 measurements were performed also for neutrons of shorter wave-lengths obtained by Bragg reflection on a fluorite crystal.

In Table II the total cross sections of BeO-filtered neutrons observed for various gases at room temperature are collected. In the third column of the table is given the sum of the total cross sections of the constituent atoms. One notices considerable differences between this sum and the observed cross sections. Except for the case of hydrogen which will be discussed later these differences are explained only to a small extent by the fact that the gas molecules are not at rest (so-called Doppler effect), and are due mostly to the interference of the waves scattered by the constituent atoms. In column 4 the values of σ calculated from the classical interference theory (see above) are given. For diatomic molecules containing two different atoms one obtains two different results, depending on the assumption that is made for the sign of the scattering lengths of the two atoms. In the cases of CO_2 and CF_4 a reasonably good agreement between theory and observation is obtained only if the scattering lengths of the component atoms are assumed to have the same sign. The same seems to be true for N_2O, though the sensitivity

TABLE II. Cross sections for $\lambda = 5.1$A (gas at room temperature).

Molecule	σ (observed)	Sum of the total cross sections of constituent atoms	σ calculated from classical interference theory
CO_2	24.5	13.0	24.8 or 4.1 for opposite phase
N_2O	57.8	34	55 or 41 for opposite phase
O_2	16.2	8.2	13.2
N_2	47.4	30	44.4
CF_4	41.5	21	38 or 7.5 for opposite phase
H_2	170	(42)	

[3] Teller and Schwinger, Phys. Rev. 52, 286 (1937).

TABLE III. Cross section of CF_4 (room temperature).

$\lambda \times 10^8$	Velocity of neutrons (m/sec.)	$\sigma \times 10^{24}$
.881	4590	19.5
1.099	3600	19.0
1.316	3000	21.5
1.492	2650	20.8
5.1	775	41.5

to a change in relative sign of the scattering length is not so pronounced here.

The very large cross section observed for H_2 is strongly perturbed by the Doppler effect because here the thermal agitation velocity of the molecules is larger than the velocity of the neutrons, so that the phenomenon can be better described by saying that the molecules hit the neutron rather than the opposite. This fact is brought out by the strong temperature dependence of the cross section which drops from 170×10^{-24} cm^2 to 81×10^{-24} cm^2 when the H_2 is cooled from room temperature to liquid air temperature. This point will be discussed later.

Table III gives the cross sections observed for CF_4 at room temperature for neutrons of various velocities. The table shows a rise by about a factor 2 when the wave-length increases from 1.5 to 5.1A. Up to 1.5A the cross section is fairly close to 21×10^{-24} cm^2, namely to the sum of the cross sections of one atom of carbon and four of fluorine. The rise is due in large measure to interference.

Tables IV and V give the results obtained on H_2 at room temperature and at liquid air temperature respectively. At room temperature one notices a steady rise in cross section with decreasing velocity of the neutrons. As previously indicated, this is to a large extent due to the thermal agitation velocity of the gas molecules. A correction factor, calculated on the assumption that the cross section is constant, is given in the fourth column. The last column gives the cross section corrected by this factor.

The data of Table V taken at liquid air temperature can be most easily compared with the Teller-Schwinger theory because at this temperature practically only rotational states 0 and 1 are involved. Column 3 of this table gives the observed cross section and column 4 gives the cross section corrected for the thermal agitation velocity of the gas molecules. The fifth column gives the theoretical values calculated with the Teller-Schwinger theory, assuming $a_0=-2.40\times10^{-12}$ and $a_1=0.54\times10^{-12}$ for the singlet and triplet state of the neutron proton system. The agreement is not too good especially for the lowest velocity, though the last value may be vitiated by the large Doppler correction.

By performing a similar set of measurements on deuterium one might be able to draw some conclusions of the spin dependence of the cross section, and we hope to make such measurements in the future.

TABLE IV. Cross section of H_2 at room temperature.

$\lambda \times 10^8$	Velocity (m/sec.)	$\sigma \times 10^{24}$	Factor for Maxwell distribution	σ/factor
.76	5212	50.9	1.048	48.6
.93	4256	53.2	1.071	49.7
1.11	3546	56.2	1.099	51.1
1.46	2714	61.2	1.172	52.2
1.68	2347	64.4	1.231	52.3
2.24	1765	68.0	1.407	48.3
5.1	775	169.7	2.514	67.5

TABLE V. Cross section of H_2 at 83°K (ortho-para ratio unchanged from room temperature).

$\lambda \times 10^8$	Velocity (m/sec.)	$\sigma_{total} \times 10^{24}$ observed	$\sigma_{Sc} \times 10^{24}$ corrected for absorption and Doppler effect	Teller-Schwinger theory
0.836	4730	44	43	—
1.202	3290	51	49	—
1.638	2420	55	52	56
1.884	2100	56	51	57
5.1	775	81	52	80

5. Total Reflection of Neutrons of Mirrors

The total reflection of neutrons on mirrors is theoretically expected to occur at very small glancing angles for substances with a positive scattering length. The index of refraction is given by

$$n=1-\lambda^2 Na/2\pi \qquad (9)$$

where a is the scattering length taken with the proper sign, N is the density of atoms in the mirror, and λ is the wave-length. Consequently n is less than 1 (case of total reflection) if a is positive. The limiting glancing angle is then

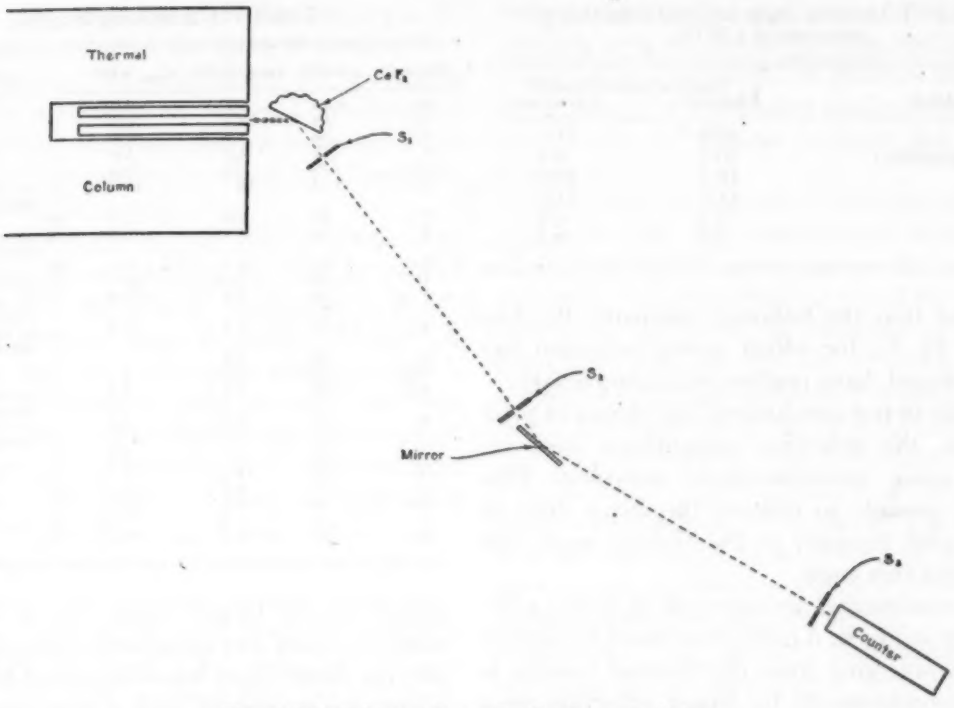


FIG. 4. Monochromatic total reflection on mirrors.

given with very good approximation by

$$\theta_0=[2(1-n)]^{\frac{1}{2}}=\lambda(Na/\pi)^{\frac{1}{2}}. \qquad (10)$$

In most cases $(n-1)$ is of the order of magnitude of 10^{-6} and the limiting angle of the order of 10′. The total reflection of neutrons on mirrors has been previously observed[4] using non-monochromatized thermal neutrons. In this case of course one cannot observe a limiting angle because θ_0 [see (10)] is proportional to λ so that neutrons of the various wave-lengths of the Maxwell distribution drop out of the reflected beam gradually as the glancing angle is increased. In spite of this, the fact that strong reflection is observed on a substance can be construed as good evidence that the index of refraction for that substance is less than one, and that therefore a is positive. This is so because, although a substance which is not totally reflecting has a finite reflection coefficient, it is so small that no very prominent reflection could be observed in such cases. From the quoted experiments it could therefore be

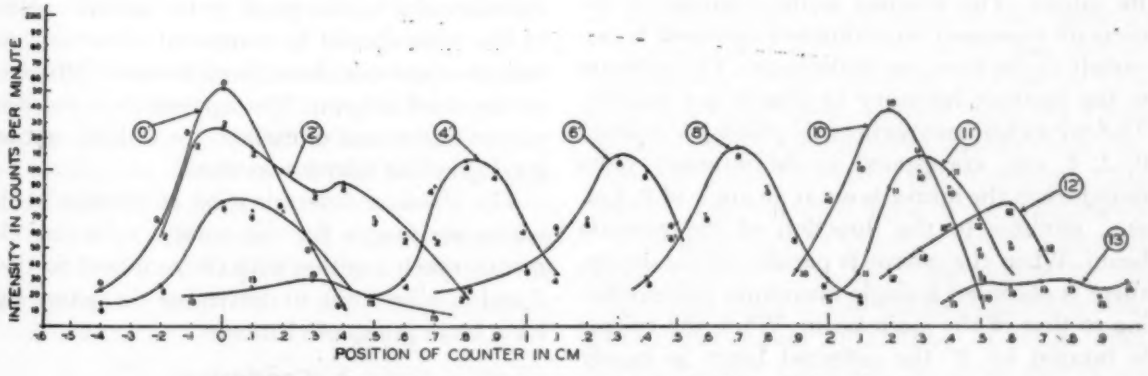


FIG. 5. Typical results of measurements of a reflection on a Be mirror.

[4] Fermi and Zinn, Phys. Rev. **70**, 103 (1946).

TABLE VI. Limiting angle for total reflection of neutrons of 1.873A.

Mirror	Limiting angle (minutes) Observed	Calculated
Be	12.0	11.1
C (graphite)	10.5	8.4
Fe	10.7	10.0
Ni	11.5	11.8
Zn	7.1	6.9
Cu	9.5	9.5

concluded that the following elements, Be, Cu, Zn, Ni, Fe, C, for which strong reflection has been observed, have positive scattering length.

In order to test conclusively the theory of total reflection, the reflection experiments were repeated using monochromatic neutrons. This made it possible to observe the sharp drop of the reflected intensity at the limiting angle and to measure this angle.

The experimental arrangement is shown schematically in Fig. 4. A collimated beam of thermal neutrons emerging from the thermal column is made monochromatic by Bragg reflection on a fluorite crystal. It is further collimated by passing through two very narrow cadmium slits, S_1 and S_2, 2 mm wide and about 3.5 meters apart. The narrow monochromatic beam emerging from the second slit falls on the mirror which can be rotated by small angles. The reflected beam is detected by a BF_3 proportional counter about 3.5 meters away and with a cadmium entrance slit 2 mm wide. By moving the counter across the beam, both direct and reflected beam are detected.

Typical results of these measurements are plotted in Fig. 5 for the case of a reflection on a Be mirror. The abscissa is the position of the detector expressed in centimeters measured transversely to the direction of the beam. The ordinate is the neutron intensity in counts per minute. The curves whose experimental points are marked 0, 2, 4, etc., correspond to the intensity plots found when the mirror is set at an angle of 0, 2, 4, etc., minutes to the direction of the neutron beam. When the mirror is parallel to the beam, there is observed a single maximum corresponding to that of the main beam. When the mirror is rotated by 2′ the reflected beam is barely resolved from the direct beam. At 4′ the two maxima are well separated. For larger angles the direct beam has not been drawn in the plot except for the largest angle, 13′, at which the reflected beam has completely disappeared and also the direct beam has disappeared because the mirror has turned by such a large angle that it completely cuts off the main beam. These points therefore correspond to the background, which actually was about 20 counts per minute.

By inspecting the plot one sees that the intensity of the reflected beam drops sharply to background between the positions 10 and 13 minutes. This drop is not quite sudden because the angular resolution was about two minutes. A study of the intensity shows the limiting angle to be 12 minutes.

In Table VI the values of the limiting angles of several mirrors measured for neutrons of de Broglie wave-length $\lambda = 1.873$A are given. The experimental values given in the second column of the table should be compared with the theoretical values calculated from formula (10) given in the third column. The agreement is excellent except in the case of carbon, for which a not too good graphite mirror was used.

The absolute determination of the sign of the scattering length for the totally reflecting elements, taken together with the results of Sections 2 and 4, allows one to determine the actual sign for a large group of elements.

TABLE VII. Scattering lengths.

Element	$a \times 10^{13}$	$4\pi a_{at}^2 \times 10^{24}$	$\sigma_{scat} \times 10^{24}$		Remarks
Ba	0.79	7.7	8	*C*	
Be	.89	8.0	6.1	*M*	
Br	.56	3.9	<7	*C*	
C	.67	4.8	4.8	*A*	
Ca	.79	7.5	9.5	*C*	
Cl	1.13	15	15	*A*,	no spin and isotope effect considered
Cu	.81	8.0	7.2	*M*	
F	.60	4.1	4.1	*A*,	no spin effect considered
Fe	.82	8.1	9.2	*M*	
H	−.39	.48	21		From theory
I	.36	1.6	1.6	*A*,	no spin effect considered
K	.35	1.5	1.5	*A*,	no spin and isotope effect considered
Li	−.59	3.4	—	*C*	
Mg	.56	3.7	4	*C*	
Mn	−.44	2.4	2.4	*A*,	no spin effect considered
N	.87	8.3	8.3	*A*,	no spin effect considered
Na	.56	3.7	3.5	*C*	
Ni	1.09	14	13	*M*	
O	.61	4.1	4.1	*A*	
Pb	.48	2.9	10	*C*	
S	.28	.95	1.1	*C*	
Zn	.58	4.1	3.6	*M*	

6. Conclusions

In Table VII all the data available at present on scattering length have been summarized. The

absolute sign given in the table has been obtained primarily from the data of Section 5 on total reflection, and indirectly, by combining these data with those of Section 2 on intensities of Bragg reflections, and Section 4, on molecular scattering. The absolute values have been obtained in various ways.

For the elements having only one isotope or one strongly dominant isotope, and no nuclear spin, the absolute value of the scattering length can be calculated from formulae (5) and (2) provided the total scattering cross section is known. For example, this method has been applied in assigning the value 0.61×10^{-12} cm for the scattering length of oxygen in a crystal. This differs by a factor 16/17 from the scattering length 0.57×10^{-12} cm for a free oxygen atom. In some cases this same method has been applied to elements having several important isotopes, or a spin, or both. In such cases the absolute value is naturally much less reliable. The method has been used when no better way was available. In the last column of Table 7 the code letter *A* has been used to denote values obtained by this method.

The results of Section 2 allow the scattering lengths of several elements to be related directly or indirectly to that of oxygen whose scattering length 0.61×10^{-12} has been used as a standard for these cases.

For the five elements Na, K, Cl, Br and I, which are not conclusively related to the others (see Section 2), Cl has been taken as a standard with a scattering length of 1.13×10^{-12} cm. The sign has been taken positive partly on account of the evidence from Bragg reflection on $NaNO_3$, as discussed in Section 2. Furthermore, both theoretical and experimental evidence indicate that cases of negative scattering length are very improbable. The probability that five elements all having the same sign should be negative is therefore exceedingly small.

In other cases the value of the scattering length has been calculated from the value of the limiting angle using formula (10). Code letter *M* indicates these instances.

In the third column of Table VII is given the expression $4\pi a^2(A/A+1)^2\times10^{24}$. If there are no important isotopes and no spin this should coincide with the scattering cross section. Otherwise it should in general be smaller, the more so the greater the differences are in the scattering lengths for the various isotopes and spin orientations.

The values in column 3 therefore should be compared with the experimental values of the scattering cross section given in column 4. Naturally the comparison is trivial for the cases designated by *A*, where formulae (2) and (5) have been used to calculate *a*. Comparison in the remaining cases shows that usually the data of the third and fourth columns are rather close. The few exceptions for which the data in column 3 are less than those of column 4 can be explained as due to experimental inaccuracy. The fact that they are close seems to indicate that, in general, the spin and isotope dependence of the scattering length is not extremely pronounced.

There are two notable exceptions. One is hydrogen, a case well known from the Teller-Schwinger theories. The other one is lead where perhaps some of the many isotopes may have a negative scattering length.

The table shows that the scattering length is positive in the great majority of cases as discussed in Section 1.

This work was performed at the Argonne National Laboratory, and we greatly benefited by the help of the staff of the laboratory in providing and operating the neutron source, and in supplying many of the experimental facilities needed. Several of the crystals used were loaned to us by the Field Museum in Chicago through the courtesy of Mr. Changnon, by Dr. Howland of the Minerology Department of Northwestern University, and by Dr. Fisher of the Geology Department of the University of Chicago. Also some specimens were obtained from the Union Carbide and Carbon Research Laboratories, Inc., and the Eagle Picher Lead Company. Mr. Warren Nyer helped in some of the experiments.

This document is based on work performed under Contract No. W-31-109-eng-38 for the Manhattan Project, and the information covered therein will appear in Division IV of the Manhattan Project Technical Series as part of the contribution of the Argonne National Laboratory.

Neutron Scattering in Ortho- and Parahydrogen

MORTON HAMERMESH
New York University, University Heights, New York, New York

AND

JULIAN SCHWINGER
Harvard University, Cambridge, Massachusetts

The cross section for the scattering of cold neutrons by parahydrogen is sensitive to the range of the forces in the triplet state. To facilitate the determination of the range from such measurements, expressions have been obtained for the dependence of the ortho and para cross sections on neutron energy and hydrogen temperature. The interpretation of transmission measurements is complicated by the existence of radiative capture, which is comparable in importance to scattering in parahydrogen. The different energy dependence of the two processes should permit an accurate determination of the individual cross sections.

SEVERAL papers[1] have emphasized the importance of measurements of neutron scattering in gaseous ortho- and parahydrogen, as a means for obtaining accurate information concerning the range of nuclear forces. Preliminary experiments were performed by Alvarez and Pitzer,[2] but with insufficient precision for this purpose.* In anticipation of more accurate experiments, a letter to *The Physical Review*[3] presents curves showing the dependence of the ortho- and parahydrogen cross sections on the range of the nuclear forces and the free-proton cross section. These curves must be corrected for three reasons:

1. The calculation of the triplet amplitude, a_1, was based on the formula: $a_1=-(1+\alpha r_0)^{\frac{1}{2}}/\alpha$, where $\alpha=[ME_0/\hbar^2]^{\frac{1}{2}}$, E_0 is the binding energy of the deuteron, and r_0 is the range of a rectangular potential well interaction. This expression is an approximate one, whose validity is restricted to ranges that are small in comparison with $1/\alpha=4.35\times10^{-13}$ cm. This approximation may lead to errors of as much as 1×10^{-13} cm in the range deduced from the observed para scattering cross section.

2. The curves are based on equations given by Schwinger[4] for the cross sections at a hydrogen gas temperature $T=20°$K, and neutron energy $E=kT$. The constants in these equations were obtained by an approximate procedure and are in error by a few percent.

3. The underlying theory has been based on a rigid-rotator model for the H_2 molecule. This approximation introduces small but not negligible errors.**

Accurate calculations of σ_{para} and σ_{ortho} have been made using essentially the same procedure as for D_2,[5] but modified to include the effect of zero-point oscillations. The only change in the theory is the replacement of $j_L(kr_e)$ by $\langle j_L(kr)\rangle_{\text{Av}}$, the mean value of $j_L(kr)$ in the lowest vibrational state. In the calculation of this average we have performed an expansion in powers of $y=r-r_e/r_e$ and dropped terms beyond $\bar{y}$ and $\langle y^2\rangle_{\text{Av}}$. It should be noted that $\bar{y}$ must be included since the zero-point vibrations are anharmonic. $\langle y^3\rangle_{\text{Av}}$ is obtained directly from the harmonic oscillator model which gives $\langle y^2\rangle_{\text{Av}}=\frac{1}{2}E_1/\hbar\omega$ where $\hbar\omega(=0.513$ ev) is the excitation energy of the first vibrational state, and $E_1(=0.0147$ ev) is the energy of the $J=1$ rotational level of H_2. A numerical value of $\bar{y}$ can be obtained by considering the dependence of the rotational energy levels on the vibration quantum number. One need only compare $r_e=0.7414\times10^{-8}$ cm, the equilibrium separation, with $r_0=0.7506$

[1] J. Schwinger and E. Teller, Phys. Rev. **52**, 286 (1937); J. Schwinger, Phys. Rev. **58**, 1004 (1940).
[2] L. W. Alvarez and K. S. Pitzer, Phys. Rev. **58**, 1003 (1940).
* Recent experiments at Los Alamos have shown that ortho-para conversion during the course of the experiments is a serious source of error. The data of these investigators, which differs considerably from that of Alvarez and Pitzer, will be published shortly.
[3] C. S. Wu, L. J. Rainwater, W. W. Havens, Jr., and J. R. Dunning, Phys. Rev. **69**, 236 (1946).
[4] J. Schwinger, Phys. Rev. **58**, 1004 (1940).
** We are indebted to E. Teller for discussions of this point.
[5] M. Hamermesh and J. Schwinger, Phys. Rev. **69**, 145 (1946).

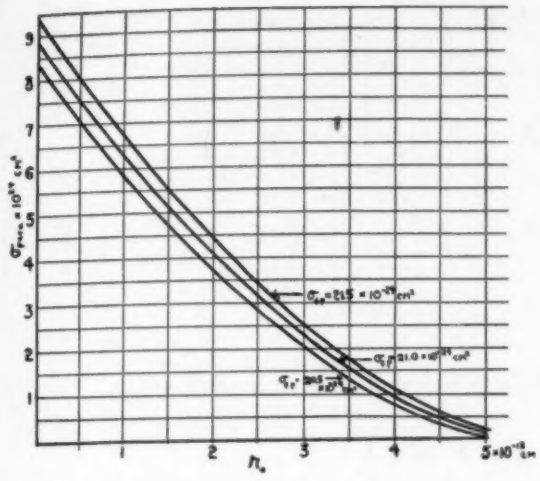


FIG. 1. Dependence of parahydrogen cross section on the range of nuclear forces. (Gas temperature $T=20°K$, neutron energy $E=kT$.)

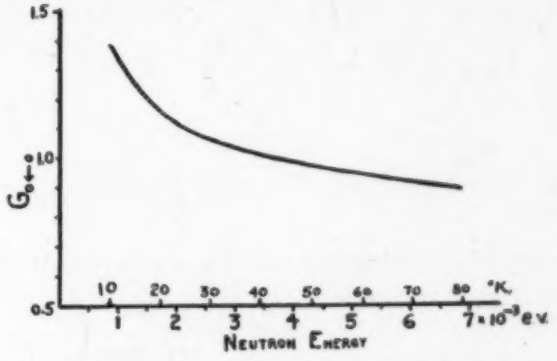


FIG. 3. Parahydrogen cross section as a function of neutron energy, omitting amplitude factors (cf. Eq. (1)). (Gas temperature = 20°K.)

$\times 10^{-8}$ cm, which is defined by $1/r_0^2=\langle 1/r^2\rangle_{Av}\simeq 1/r_e^2(1-2\bar{y}+3\langle y^2\rangle_{Av})$. This procedure gives $\bar{y}=0.0337$, $\langle y^2\rangle_{Av}=0.0143$.

The cross sections for the various transitions have been expanded in powers of E/E_1, where E is the neutron energy. The thermal averages of the cross sections for the various transitions are:

$$\sigma_{0\leftarrow 0}=\frac{16\pi}{9}(3a_1+a_0)^2G_{0\leftarrow 0},$$

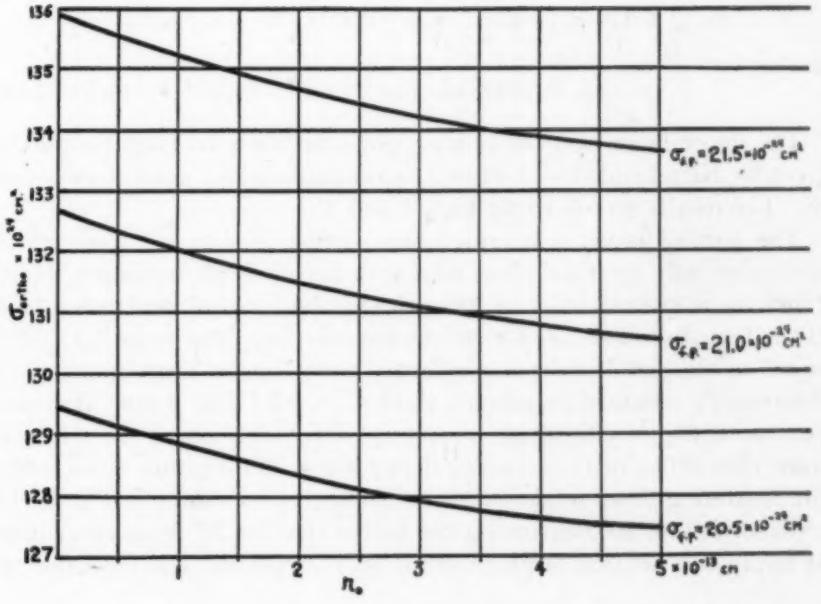


FIG. 2. Dependence of orthohydrogen cross section on the range of nuclear forces. (Gas temperature $T=20°K$, neutron energy $E=kT$.)

$$G_{0\leftarrow 0}=\frac{1}{\pi^{\frac{1}{2}}}\frac{\exp(-x^2)}{x}+\left(1+\frac{1}{2x^2}\right)\Phi(x)-\frac{8}{27}(1+2\bar{y}+\langle y^2\rangle_{Av})\frac{E}{E_1}\left[\frac{1}{\pi^{\frac{1}{2}}}\left(1+\frac{5}{2x^2}\right)\frac{\exp(-x^2)}{x}\right.$$

$$\left.+\left(1+\frac{3}{x^2}+\frac{3}{4x^4}\right)\Phi(x)\right]+\frac{512}{(81)(135)}\left(1+4\bar{y}+\frac{5}{2}\{\bar{y}\}^2+\frac{7}{2}\langle y^2\rangle_{Av}\right)\left(\frac{E}{E_1}\right)^2$$

$$\times\left[\frac{1}{\pi^{\frac{1}{2}}}\left(1+\frac{7}{x^2}+\frac{33}{4x^4}\right)\frac{\exp(-x^2)}{x}+\left(1+\frac{15}{2x^2}+\frac{54}{4x^4}+\frac{15}{8x^6}\right)\Phi(x)\right],$$

$$\sigma_{1\leftarrow 1}=\frac{16\pi}{9}[(3a_1+a_0)^2+2(a_1-a_0)^2]G_{1\leftarrow 1},$$

where $G_{1\leftarrow 1}$ is identical with $G_{0\leftarrow 0}$, except that the factor preceding $(E/E_1)^2$ is replaced by

$$\frac{1024}{(81)(225)}\left(1+4\bar{y}+\frac{11}{4}\{\bar{y}\}^2+\frac{13}{4}\langle y^2\rangle_{Av}\right);$$

$$\sigma_{0\leftarrow 1}=\frac{64\pi}{9}(a_1-a_0)^2\left(\frac{E_1}{E}\right)^{\frac{1}{2}}\left[0.01982(1+1.73\bar{y}-0.15\langle y^2\rangle_{Av})\right.$$

$$\left.+0.02719(1+1.23\bar{y}-0.664\langle y^2\rangle_{Av})\frac{E}{E_1}\left(1+\frac{3kT}{4E}\right)+0.000264\left(\frac{E}{E_1}\right)^2\left\{1+\frac{5}{2}\frac{kT}{E}+\frac{15}{16}\left(\frac{kT}{E}\right)^2\right\}\right], \quad (1)$$

$a_{1,0}$ are the triplet and singlet amplitudes, T is the temperature of the gas in °K, $x^2=2E/kT$ and

$$\Phi(x)=\frac{2}{\pi^{\frac{1}{2}}}\int_0^x\exp(-t^2)dt.$$

For $T=20$°K and $E=kT$ we obtain:

$$\sigma_{0\leftarrow 0}=6.444(3a_1+a_0)^2,\quad \sigma_{1\leftarrow 1}=6.450[(3a_1+a_0)^2+2(a_1-a_0)^2],\quad \sigma_{0\leftarrow 1}=1.753(a_1-a_0)^2,$$

so that

$$\sigma_{\text{para}}=6.444(3a_1+a_0)^2,\quad \sigma_{\text{ortho}}=6.450(3a_1+a_0)^2+14.653(a_1-a_0)^2.$$

The dependence of a_1 on r_0 was computed for a rectangular potential well, using the expansion given by Kittel and Breit.[6] Then a_0 was calculated for several values of the free-proton cross section, $\sigma_{t.p.}$ The results are shown in Figs. 1 and 2.

The parahydrogen scattering cross section is obtained from transmission measurements which determine only the total cross section of the molecule including the radiative capture cross section. From the measured value of the mean lifetime of slow neutrons in water,[7] the capture cross section of the hydrogen molecule for 20° neutrons is calculated to be 2.5×10^{-24} cm², which is approximately equal to the parahydrogen scattering cross section corresponding to the range of nuclear forces customarily assumed in current theory, ($r_0=2.8\times10^{-13}$ cm). Transmission experiments at several neutron energies will enable a separate determination of the scattering and capture cross sections since they differ in their energy dependence. The capture cross section follows a $1/v$ law, whereas the scattering cross section varies with energy as shown in Fig. 3. Thus, for example, the capture cross section for 80° neutrons is one-half of that for 20° neutrons, whereas the scattering cross section at the higher neutron temperature is only 25 percent less than that at 20°.

[6] C. Kittel and G. Breit, Phys. Rev. 56, 744 (1939).
[7] J. H. Manley, L. J. Haworth, and E. A. Luebke, Phys. Rev. 61, 152 (1942).

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

The Double Focusing Beta-Ray Spectrometer

FRANKLIN B. SHULL AND DAVID M. DENNISON
Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan
(Received February 10, 1947)

The double focusing spectrometer recently proposed by Siegbahn and Svartholm has been analyzed and compared with the conventional semicircular spectrometer. A shaped magnetic field is required which possesses cylindrical symmetry as well as symmetry with respect to the median plane of the instrument. The magnetic field in the median plane is assumed to have the form

$$H=H_0-(r-a)\alpha H_0/a+(r-a)^2\beta H_0/a^2,$$

where α and β are experimentally disposable shaping constants. The analysis, which has been carried through the second order of approximation, shows that double focusing occurs when $\alpha=\frac{1}{2}$. In general there will exist, for all types of spectrometers, second order defocusing terms which tend to lower the resolving power. However, these may be almost completely eliminated in the double focusing spectrometer by choosing $\beta=\frac{1}{2}$. A figure is presented showing the form of the focused image for a semicircular spectrometer and for several cases of the double focusing spectrometer. It becomes evident that when $\alpha=\frac{1}{2}$ and $\beta=\frac{1}{2}$, the intensity of the image as well as the resolving power become many times greater than for the former type of instrument.

1. INTRODUCTION

IT is customary in beta-ray spectroscopy to focus electrons with either the magnetic semicircular spectrometer or the magnetic electron lens. The former requires a uniform field, which is created by an iron-core magnetic circuit with plane parallel pole faces. Large magnetic field strengths are feasible, and hence high energy electrons can easily be focused. The electron lens uses a magnetic circuit which is relatively free of iron. Much copper and power are necessary to produce large fields, and it is impractical to work with high energy electrons. In another respect, however, the electron lens seems superior. Assuming equal resolving power for both, the electron lens can focus a beam intensity as much as 15 to 20 times greater than the semicircular spectrometer.

Siegbahn and Svartholm[1] have proposed and constructed an electron focusing system which utilizes an inhomogeneous magnetic field produced by an iron-core magnetic circuit with shaped pole faces. The magnetic field is shaped to fall off radially as $1/r^{\frac{1}{2}}$. The electron beam is focused after sweeping out an angle $2^{\frac{1}{2}}\pi$ about the axis of symmetry. The focusing principle is basically a refinement of the semicircular system which permits a considerable increase of beam intensity without sacrificing resolving power. Such a spectrometer could compete more favorably with the electron lens in terms of focused beam intensity, while retaining the ability to focus high energy particles.

It is proposed in the paper which follows to analyze this focusing system, evaluate its performance, and compare it to the conventional semicircular type. The equations of motion of the electrons will be solved by means of the usual perturbation methods, and the calculations will be carried through the second order of approximation.

2. THEORY

It will be convenient to use the cylindrical coordinates r, θ, and z, and it will be assumed that the magnetic field is (a) independent of the azimuth angle θ and (b) symmetrical with respect to the median plane $z=0$. The vertical component of the field, H_z, as measured in the median plane, is taken to have the form

$$H_z=H_0-\left(\frac{r-a}{a}\right)\alpha H_0+\left(\frac{r-a}{a}\right)^2\beta H_0-\cdots, \quad (1)$$

where a is the radial coordinate of the center of the electron source (whose other coordinates are $\theta=0$, $z=0$), H_0 is a constant magnetic field, and α and β are arbitrary dimensionless constants. Clearly Eq. (1) represents a series development for H_z through second order, and assumes that the equilibrium or zeroth order orbit will be a circle with the radius $r=a$. If, for points in the

[1] Siegbahn and Svartholm, Nature **157**, 872 (1946).

median plane, H_z is assumed to fall off as $1/r^n$, the identifications $\alpha=n$ and $\beta=(n^2+n)/2$ can be made, but it should be emphasized that H_0, α, and β are experimentally disposable parameters in the present problem. Since curl$H=0$, it follows that the radial component H_r of the magnetic field must be

$$H_r=zH_0\left[-\frac{\alpha}{a}+\frac{2\beta(r-a)}{a^2}\right],$$

and since div$H=0$, it follows that a fourth term must be added to Eq. (1) in order to express H_z at points off the median plane. Thus,

$$H_z=H_0-\left(\frac{r-a}{a}\right)\alpha H_0+\left(\frac{r-a}{a}\right)^2\beta H_0-\frac{z^2}{a^2}\beta H_0.$$

In any purely magnetic field, the motion of a charged particle must be such that its velocity, and consequently its mass, will remain constant. The equations of motion for the particle then reduce to the following expressions:

$$\frac{d\dot{r}}{dt}-r\dot{\theta}^2=\frac{e}{mc}r\dot{\theta}H_z, \tag{2}$$

$$\frac{d}{dt}(r^2\dot{\theta})=\frac{er}{mc}(\dot{z}H_r-\dot{r}H_z), \tag{3}$$

$$\frac{d\dot{z}}{dt}=-\frac{e}{mc}r\dot{\theta}H_r. \tag{4}$$

As is well known, an integral of the motion may be obtained by multiplying (2) by $\dot{r}$, (3) by $\dot{\theta}$, and (4) by $\dot{z}$, summing, and integrating with respect to the time. This integral is

$$\dot{r}^2+r^2\dot{\theta}^2+\dot{z}^2=v^2=\text{constant} \tag{5}$$

where v is the initial velocity of the electron. Any three of the Eqs. (2), (3), (4), and (5) would be sufficient to describe the motion, and it will prove convenient in the present case to eliminate (3) and use only (2), (4), and (5).

Assume that an electron originates in the source at the point $r=a$, $z=0$, $\theta=0$ and has the initial velocity components $\dot{r}=0$, $\dot{z}=0$, $r\dot{\theta}=v=-eH_0a/mc$. Under these conditions, the solution of the equations of motion becomes

$$r=a, \quad z=0, \quad \theta=\omega t,$$

where $\omega=v/a$. This circular orbit is hereafter designated as the equilibrium or zeroth order orbit. In general, however, the initial conditions describing an electron orbit will be

$$\left.\begin{aligned} r&=a+\delta r\\ z&=\delta z\\ \theta&=0\\ \dot{r}/v&=\varphi_r\\ \dot{z}/v&=\varphi_z \end{aligned}\right\} t=0,$$

where φ_r and φ_z are evidently angles which specify the direction of the initial velocity, and δr and δz are coordinates of the initial position relative to the center of the source.

It is proposed to solve Eqs. (2), (4), and (5) by standard perturbation methods, and to obtain solutions for r, θ, and z which are complete through second order. Let

$$\begin{aligned} r&=a+\lambda\rho_1+\lambda^2\rho_2,\\ \dot{\theta}&=\omega+\lambda\theta_1+\lambda^2\theta_2,\\ z&=\quad\lambda\zeta_1+\lambda^2\zeta_2, \end{aligned} \tag{6}$$

where ρ_1, θ_1, ζ_1 and ρ_2, θ_2, ζ_2 are functions of the time, and λ is a parameter of smallness. It will be noticed that the equations of motion involve $\dot{\theta}$ rather than θ, and consequently $\dot{\theta}$ has been specified above in place of θ. The variables (6) are substituted directly into Eqs. (2), (4), and (5), and the three resulting equations are classified according to powers of the parameter λ. By equating the coefficients of λ^0, λ^1, and λ^2 to zero, equations are obtained which describe the equilibrium orbit and the first- and second-order corrections to the equilibrium orbit respectively.

The zeroth-order equations yield the following relations, which determine the magnetic field H_0 as a function of the momentum of the electron.

$$v=\omega a, \quad \omega=-eH_0/mc. \tag{7}$$

The first-order equations are

$$\begin{aligned} &\lambda\frac{d\dot{\rho}_1}{dt}+\lambda\omega^2(1-\alpha)\rho_1=0,\\ &\lambda\frac{d\dot{\zeta}_1}{dt}+\lambda\omega^2\alpha\zeta_1=0,\\ &\theta_1=-\omega\rho_1/a. \end{aligned} \tag{8}$$

It is evident that the first-order solutions $\lambda\rho_1$, $\lambda\theta_1$, and $\lambda\zeta_1$ will be sinusoidal, with circular frequencies $\omega(1-\alpha)^{\frac{1}{2}}$, $\omega(1-\alpha)^{\frac{1}{2}}$, and $\omega\alpha^{\frac{1}{2}}$, respectively.

The second-order equations become after some simplification

$$\lambda^2\left[\frac{d\dot\rho_2}{dt}+\omega^2(1-\alpha)\rho_2\right]=\lambda^2\left[\frac{\omega^2}{a}\rho_1^2-\frac{\beta\omega^2}{a}(\rho_1^2-\zeta_1^2)-\frac{1}{2a}(\dot\rho_1^2+\dot\zeta_1^2)\right], \tag{9}$$

$$\lambda^2\left[\frac{d\dot\zeta_2}{dt}+\omega^2\alpha\zeta_2\right]=2\lambda^2\frac{\beta\omega^2}{a}\rho_1\zeta_1, \tag{10}$$

$$\lambda^2\theta_2=\lambda^2\left[\frac{\omega}{a^2}\rho_1^2-\frac{\omega}{a}\rho_2-\frac{(\dot\rho_1^2+\dot\zeta_1^2)}{2a^2\omega}\right]. \tag{11}$$

Both Eqs. (9) and (10) consist of a homogeneous part, to the left of the equality sign, and an inhomogeneous part. The solution of the homogeneous part of each may be neglected, since it is identical to the solution of the corresponding first order Eq. (8), and thus contributes nothing new. The inhomogeneous parts which remain are now known functions of the time, and may be computed from the first order solutions. The equations can be integrated by elementary processes.

The final expressions for r, z, and θ, complete through second order, are

$$\begin{aligned}
r=a&+\left[\delta r+\frac{\delta r^2}{a}\frac{(\beta-\alpha)}{3(1-\alpha)}+a\varphi_r^2\frac{(4\beta-\alpha-3)}{6(1-\alpha)^2}+\frac{\delta z^2}{a}\frac{(3\alpha\beta-\beta-\alpha^2)}{(1-\alpha)(1-5\alpha)}+a\varphi_z^2\frac{(4\beta-3\alpha+1)}{2(1-\alpha)(1-5\alpha)}\right]\cos\omega(1-\alpha)^{\frac12}t\\
&+\frac{1}{(1-\alpha)^{\frac12}}\left[a\varphi_r-\frac{(\alpha+2\beta-3)}{3(1-\alpha)}\varphi_r\delta r-\frac{(\alpha+2\beta)}{(1-5\alpha)}\varphi_z\delta z\right]\sin\omega(1-\alpha)^{\frac12}t\\
&+\frac{1}{4a(1-\alpha)}\left[(\alpha-2\beta+1)\left(\delta r^2+\frac{a^2\varphi_r^2}{1-\alpha}\right)+(2\beta-\alpha)\left(\delta z^2+\frac{a^2\varphi_z^2}{\alpha}\right)\right]\\
&+\frac{(\alpha+2\beta-3)}{12a(1-\alpha)}\left(\delta r^2-\frac{a^2\varphi_r^2}{1-\alpha}\right)\cos2\omega(1-\alpha)^{\frac12}t+\frac{(\alpha+2\beta)}{4a(1-5\alpha)}\left(\delta z^2-\frac{a^2\varphi_z^2}{\alpha}\right)\cos2\omega\alpha^{\frac12}t\\
&+\frac{(\alpha+2\beta-3)}{6(1-\alpha)^{\frac12}}\varphi_r\delta r\sin2\omega(1-\alpha)^{\frac12}t+\frac{(\alpha+2\beta)}{2(1-5\alpha)\alpha^{\frac12}}\varphi_z\delta_z\sin2\omega\alpha^{\frac12}t,
\end{aligned} \tag{12}$$

$$\begin{aligned}
z=&\left[\delta z-\frac{2\beta}{a}\left(\frac{\delta r\delta z}{5\alpha-1}-\frac{2a^2\varphi_r\varphi_z}{5\alpha^2-6\alpha+1}\right)\right]\cos\omega\alpha^{\frac12}t\\
&+\frac{1}{\alpha^{\frac12}}\left[a\varphi_z-2\beta\left(\frac{\varphi_z\delta r}{1-5\alpha}+\frac{(3\alpha-1)}{5\alpha^2-6\alpha+1}\varphi_r\delta z\right)\right]\sin\omega\alpha^{\frac12}t\\
&+\frac{\beta\sin\omega[(1-\alpha)^{\frac12}+\alpha^{\frac12}]t}{\alpha-1-2(\alpha-\alpha^2)^{\frac12}}\left[\frac{\varphi_r\delta z}{(1-\alpha)^{\frac12}}+\frac{\varphi_z\delta r}{\alpha^{\frac12}}\right]+\frac{\beta\sin\omega[(1-\alpha)^{\frac12}-\alpha^{\frac12}]t}{\alpha-1+2(\alpha-\alpha^2)^{\frac12}}\left[\frac{\varphi_r\delta z}{(1-\alpha)^{\frac12}}-\frac{\varphi_z\delta r}{\alpha^{\frac12}}\right]\\
&+\frac{\beta\cos\omega[(1-\alpha)^{\frac12}+\alpha^{\frac12}]t}{\alpha-1-2(\alpha-\alpha^2)^{\frac12}}\left[\frac{\delta r\delta z}{a}-\frac{a\varphi_r\varphi_z}{(\alpha-\alpha^2)^{\frac12}}\right]+\frac{\beta\cos\omega[(1-\alpha)^{\frac12}-\alpha^{\frac12}]t}{\alpha-1+2(\alpha-\alpha^2)^{\frac12}}\left[\frac{\delta r\delta z}{a}+\frac{a\varphi_r\varphi_z}{(\alpha-\alpha^2)^{\frac12}}\right],
\end{aligned} \tag{13}$$

$$\theta=\left[1+\frac{(2\beta-\alpha-\alpha^2)}{4(1-\alpha)}\left(\frac{\delta r^2}{a^2}+\frac{\varphi_r^2}{1-\alpha}\right)+\frac{(\alpha^2-2\beta)}{4(1-\alpha)}\left(\frac{\delta z^2}{a^2}+\frac{\varphi_z^2}{\alpha}\right)\right]\omega t$$
$$-\frac{1}{(1-\alpha)^{\frac{1}{2}}}\left[\frac{\delta r}{a}+\frac{\delta r^2}{a^2}\frac{(\beta-\alpha)}{3(1-\alpha)}+\varphi_r^2\frac{(4\beta-\alpha-3)}{6(1-\alpha)^2}+\frac{\delta z^2}{a^2}\frac{(3\alpha\beta-\beta-\alpha^2)}{(1-\alpha)(1-5\alpha)}\right.$$
$$\left.+\varphi_z^2\frac{(4\beta-3\alpha+1)}{2(1-\alpha)(1-5\alpha)}\right]\sin\omega(1-\alpha)^{\frac{1}{2}}t+\frac{1}{1-\alpha}\left[\varphi_r-\frac{\varphi_r\delta r}{3a}\frac{(\alpha+2\beta-3)}{(1-\alpha)}-\frac{\varphi_z\delta z}{a}\frac{(\alpha+2\beta)}{(1-5\alpha)}\right]\cos\omega(1-\alpha)^{\frac{1}{2}}t$$
$$+\frac{(3\alpha^2-13\alpha-2\beta+12)}{24(1-\alpha)^4}\left[\frac{\delta r^2}{a^2}-\frac{\varphi_r^2}{1-\alpha}\right]\sin2\omega(1-\alpha)^{\frac{1}{2}}t-\frac{(5\alpha^2+2\beta)}{8\alpha^{\frac{1}{2}}(1-5\alpha)}\left[\frac{\delta z^2}{a^2}-\frac{\varphi_z^2}{\alpha}\right]\sin2\omega\alpha^{\frac{1}{2}}t$$
$$-\frac{(3\alpha^2-13\alpha-2\beta+12)}{12(1-\alpha)^2}\frac{\varphi_r\delta r}{a}\cos2\omega(1-\alpha)^{\frac{1}{2}}t+\frac{(5\alpha^2+2\beta)}{4\alpha(1-5\alpha)}\frac{\varphi_z\delta z}{a}\cos2\omega\alpha^{\frac{1}{2}}t$$
$$-\frac{\varphi_r}{(1-\alpha)}+\frac{(\alpha^2-3\alpha+2\beta)}{4(1-\alpha)^2}\frac{\varphi_r\delta r}{a}-\frac{(\alpha^2+2\beta)}{4\alpha(1-\alpha)}\frac{\varphi_z\delta z}{a}. \quad (14)$$

The integration constants have been chosen so that θ, r, and z satisfy the initial conditions given earlier, namely, at $t=0$

$$r=a+\delta r, \quad \theta=0, \quad z=\delta z, \quad \dot r=a\omega\varphi_r=v\varphi_r, \quad \dot z=a\omega\varphi_z=v\varphi_z, \quad v^2=\dot r^2+r^2\dot\theta^2+\dot z^2.$$

It should be noted that in the orbital Eqs. (12), (13), and (14) those terms which do not contain δr, δz, φ_r or φ_z are the zeroth-order terms, while those which are linear and quadratic in these quantities represent the first and second order terms respectively.

3. THE SEMICIRCULAR SPECTROMETER

When $\alpha=\beta=0$, Eqs. (12), (13), and (14) describe the electron orbits in a semicircular spectrometer, and it will be fruitful to consider this case in some detail before proceeding further. Equations (12) and (13) become

$$r=a+\left[\delta r-\frac{a}{2}(\varphi_r^2-\varphi_z^2)\right]\cos\omega t+\varphi_r(a+\delta r)\sin\omega t-\frac{1}{4a}(\delta r^2-a^2\varphi_r^2)\cos2\omega t$$
$$-\tfrac{1}{2}\varphi_r\delta r\sin2\omega t+\frac{1}{4a}(\delta r^2+a^2\varphi_r^2)-\frac{a\varphi_z^2}{2}, \quad (15)$$

$$z=\delta z+a\varphi_z\omega t. \quad (16)$$

When the magnetic field is uniform, as it is here, there exists no restoring force in the z direction, and consequently there can be no z focusing. However, the first order sinusoidal terms in r show that r focusing will occur when $\omega t=\pi$, and that the image will be reversed in r with respect to the source. Let θ^*, r^*, z^* be the coordinates of a focused electron, where $\theta^*=\pi$. Let t^* be the time required for an electron to travel from the source to the focused position. In zeroth order

$$\theta=\omega t, \quad \theta^*=\omega t^*=\pi, \quad (17)$$

while in first order

$$\theta=\omega t-\frac{\delta r}{a}\sin\omega t+\varphi_r\cos\omega t-\varphi_r,$$
$$\theta^*=\pi=\omega t^*-2\varphi_r, \quad \omega t^*=\pi+2\varphi_r. \quad (18)$$

It will not be necessary for the present purpose to obtain the second-order correction to t^*. Substituting t^* from (17) and (18) into the second- and first-order terms, respectively, of (15) and (16), r^* and z^* are found to be

$$r^*=a-\delta r-a(\varphi_r^2+\varphi_z^2), \tag{19}$$

$$z^*=\delta z+\pi a\varphi_z+2a\varphi_r\varphi_z. \tag{20}$$

It is evident that the second-order terms in Eq. (19) represent a departure from perfect focusing which is strongly dependent on φ_r and φ_z, and this will be discussed later. Since the orbits are all circular in form, it is clear that rigorous expressions for r^* and z^* may be derived from purely geometric considerations. When these exact expressions are developed through second order, they yield Eqs. (19) and (20), as they must.

4. THE DOUBLE FOCUSING SPECTROMETER

When the first-order sinusoidal frequencies in Eqs. (12) and (13) are made precisely equal, a simultaneous focusing occurs in both the r and z directions. The necessary condition for this two-dimensional focusing is

$$\omega(1-\alpha)^{\frac{1}{2}}=\omega\alpha^{\frac{1}{2}}, \quad \alpha=\tfrac{1}{2}.$$

In this case, (12) and (13) become

$$r=a+\left[\delta r+\frac{(2\beta-1)}{3a}(\delta r^2-\delta z^2)+\frac{(8\beta-7)}{3}a\varphi_r^2+\frac{(1-8\beta)}{3}a\varphi_z^2\right]\cos\frac{\omega t}{2^{\frac{1}{2}}}$$

$$+2^{\frac{1}{2}}\left[a\varphi_r+\frac{(5-4\beta)}{3}\varphi_r\delta_r+\frac{(1+4\beta)}{3}\varphi_z\delta z\right]\sin\frac{\omega t}{2^{\frac{1}{2}}}+\frac{(3-4\beta)}{4a}(\delta r^2+2a^2\varphi_r^2)+\frac{(4\beta-1)}{4a}(\delta z^2+2a^2\varphi_z^2)$$

$$+\frac{1}{12a}[(4\beta-5)(\delta r^2-2a^2\varphi_r^2)-(1+4\beta)(\delta z^2-2a^2\varphi_z^2)]\cos 2^{\frac{1}{2}}\omega t$$

$$+\frac{2^{\frac{1}{2}}}{6}[(4\beta-5)\varphi_r\delta r-(1+4\beta)\varphi_z\delta z]\sin 2^{\frac{1}{2}}\omega t, \tag{21}$$

$$z=\left[\delta z-\frac{4\beta}{3a}\delta r\delta z-\frac{16\beta}{3}a\varphi_r\varphi_z\right]\cos\frac{\omega t}{2^{\frac{1}{2}}}+2^{\frac{1}{2}}\left[a\varphi_z+\frac{4\beta}{3}(\varphi_z\delta r+\varphi_r\delta z)\right]\sin\frac{\omega t}{2^{\frac{1}{2}}}$$

$$+2\beta\left[\frac{\delta r\delta z}{a}+2a\varphi_r\varphi_z\right]-\frac{2\beta}{3}\left[\frac{\delta r\delta z}{a}-2a\varphi_r\varphi_z\right]\cos 2^{\frac{1}{2}}\omega t-\frac{2^{\frac{1}{2}}\beta}{3}[\varphi_z\delta r+\varphi_r\delta z]\sin 2^{\frac{1}{2}}\omega t. \tag{22}$$

The first order sinusoidal terms in (21) and (22), namely those terms which are linear in δr, δz, φ_r or φ_z, show that focusing will take place when $\omega t/2^{\frac{1}{2}}=\pi$, and that the image will be reversed in r and inverted in z. Let θ^*, r^*, z^* and t^* be defined as before. Then, in zeroth approximation,

$$\theta=\omega t, \quad \theta^*=\omega t^*=2^{\frac{1}{2}}\pi\sim 254^\circ\,33', \tag{23}$$

while in first order

$$\theta=\omega t-2^{\frac{1}{2}}\frac{\delta r}{a}\sin\frac{\omega t}{2^{\frac{1}{2}}}+2\varphi_r\cos\frac{\omega t}{2^{\frac{1}{2}}}-2\varphi_r,$$

$$\theta^*=2^{\frac{1}{2}}\pi=\omega t^*-4\varphi_r, \quad \omega t^*=2^{\frac{1}{2}}\pi+4\varphi_r. \tag{24}$$

Substituting t^* from Eqs. (23) and (24) into the second- and first-order terms, respectively, of

(21) and (22), r^* and z^* are found to be

$$r^*=a-\delta r+\frac{(2-4\beta)}{3}\frac{(\delta r^2-\delta z^2)}{a}+\frac{(2-16\beta)}{3}a(\varphi_r^2-\varphi_z^2), \tag{25}$$

$$z^*=-\delta z+\frac{8\beta}{3}\frac{\delta r\delta z}{a}+\frac{(32\beta-12)}{3}a\varphi_r\varphi_z. \tag{26}$$

As before, the second-order terms in (25) and (26) represent departures from perfect focusing in r and z, respectively. The arbitrary constant β, which determines the second-order correction to the magnetic field, still remains as a disposable parameter and can be chosen so as to minimize the second order defocusing.

For the spectrometer of Siegbahn and Svartholm, it is stated that the magnetic field was shaped to fall off as $1/r^{\frac{1}{2}}$. This is identical with Eq. (1) if $\alpha=\frac{1}{2}$ and $\beta=\frac{3}{8}$. When $\beta=\frac{3}{8}$, Eqs. (25) and (26) become

$$r^*=a-\delta r+\frac{(\delta r^2-\delta z^2)}{6a}-\frac{4}{3}a(\varphi_r^2-\varphi_z^2), \tag{27}$$

$$z^*=-\delta z+\frac{\delta r\delta z}{a}. \tag{28}$$

As in the semicircular spectrometer, the second order defocusing in r is strongly dependent upon φ_r and φ_z, and this point will be discussed in greater detail later.

With a different choice of the field parameter β, it is possible to eliminate φ_r and φ_z entirely from the second order defocusing in r. Thus, when $\beta=\frac{1}{8}$, Eqs. (25) and (26) become

$$r^*=a-\delta r+\frac{(\delta r^2-\delta z^2)}{2a}, \tag{29}$$

$$z^*=-\delta z+\frac{\delta r\delta z}{3a}-\frac{8}{3}a\varphi_r\varphi_z. \tag{30}$$

Comparing Eqs. (19) and (20), (27) and (28), and (29) and (30), the following conclusions may be drawn. It will be assumed throughout that the source is rectilinear in shape, with its long dimension parallel to the z axis.

(a) For the semicircular spectrometer, the negative term $a(\varphi_r^2+\varphi_z^2)$ in Eq. (19) predicts that the center of gravity of the image will be shifted to smaller values of r, and that the image will be correspondingly broadened.

(b) For the double focusing spectrometer where $\beta=\frac{3}{8}$, the term $4a(\varphi_r^2-\varphi_z^2)/3$ in Eq. (27) may be either positive or negative. The image will be broadened toward both larger and smaller values of r. Its center of gravity may or may not shift; the magnitude and direction of the shift depend upon the relative maximum values of φ_r and φ_z. The term $(\delta r^2-\delta z^2)/6a$ in Eq. (27) predicts that the image will be slightly concave toward the origin of coordinates.

(c) For the double focusing spectrometer where $\beta=\frac{1}{8}$, Eq. (29) predicts a curvature of the image similar to that discussed under (b) above. However, the absence of φ_r and φ_z means that the image will be sharp, and its center of gravity is at $r=a$.

(d) Terms of similar order of magnitude involving $a\varphi_r\varphi_z$ appear in each of the z^* Eqs. (20) and (30), but not in (28). In addition, Eq. (20), for the semicircular spectrometer, contains the much stronger term $\pi a\varphi_z$. Such terms mean that the image will be lengthened vertically, and this effect is especially prominent in the semicircular type.

The conclusions listed under (a) through (d) above are illustrated graphically in Fig. 1. The rectilinear source has been assumed to have the dimensions 20 mm$\times$2 mm, while the radius a has been arbitrarily chosen as $a=10$ cm. It is further assumed that the angles φ_r and φ_z are limited by a system of baffles so that $(\varphi_r^2+\varphi_z^2)^{\frac{1}{2}}\leqq 0.1$ radian. Figure 1 illustrates the image formed in each of the three spectrometer types which have been discussed; only the upper half of each image is shown, since the image is symmetrical with respect to the median plane. The shaded rectangle represents a perfect image of the source, centered at $r=a$; it could also be considered as representing an exit slit, through which the focused beam passes into a Geiger

counter. The dotted contours represent the extreme boundaries of the focused image; it should be emphasized that the intensity distribution within the boundary is not uniform. Nevertheless, it is evident that a large proportion of the focused particles in a semicircular spectrometer miss the exit slit entirely, and consequently the counting rate may be expected to be relatively low. On the other hand, the double focusing spectrometer, especially when the field parameter is $\beta=\frac{1}{8}$, makes very efficient use of the focused beam, and the counting rate will be relatively high. Moreover, when $\beta=\frac{1}{8}$, the sharpness of the image leads to a considerably higher resolving power. If the exit slit is properly curved and elongated to fit the image, maximum counting rate and maximum resolving power will be attained simultaneously.

5. CONCLUSION

The analysis has shown that, through the use of a shaped magnetic field, the beta-ray spectrometer proposed by Siegbahn and Svartholm may be made into a very effective instrument. The resolving power as well as the intensity of the focused beam are much higher than those of a semicircular spectrometer. Some technical difficulties may be encountered in the construction of the instrument, however, since it is necessary that the shape of the magnetic field shall remain independent of the magnitude of the field over the useful range of operation.

With the shaped field where $\beta=\frac{1}{8}$ and where a curved exit slit is used, the resolving power is very high; in fact, the equations show that the image will almost precisely coincide with the slit. This nearly perfect predicted performance may be slightly illusory in the sense that third and higher order terms in the development of the solution will again lead to some defocusing. It is clear, however, that these will be small and of the order of magnitude of, say, $\delta z^2/a^2$ or of $a\varphi_r^3$. For the spectrometer dimensions chosen in the present example, these terms would have extreme values of about 0.1 mm. In principle, through the use of a further shaping constant γ for the magnetic field, it might be possible to cancel out much of the third order defocusing. This would appear to be too refined a process to be practical however.

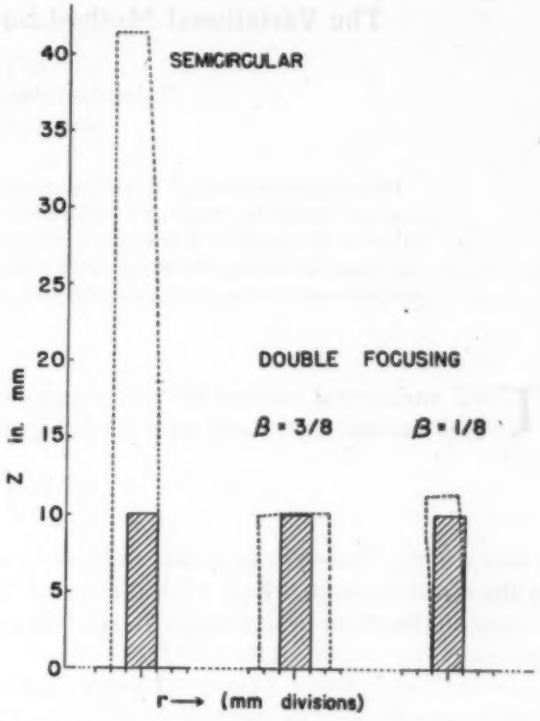


FIG. 1. The focused images for a semicircular spectrometer and for double focusing spectrometers with $\beta=\frac{3}{8}$ and $\frac{1}{8}$, respectively.

The Variational Method for Asymptotic Neutron Densities*

R. E. MARSHAK
University of Rochester, Rochester, New York

(Received January 24, 1947)

Inhomogeneous integral equations connected with a certain class of neutron diffusion problems are treated by means of a variational method. It is shown how to obtain quite accurate values of the asymptotic neutron density in the following three cases: (a) Milne's problem for the plane, (b) infinite scattering medium surrounding black spherical core, (c) infinite scattering medium surrounding black spherical core with air gap.

THE variational method for solving eigenvalue problems associated with integral equations is well known. One starts with the integral equation:

$$\rho_0(\mathbf{r}) = c\int d\mathbf{r}'\rho_0(\mathbf{r}')K(\mathbf{r},\mathbf{r}'), \tag{1}$$

where c is the (lowest) eigenvalue to be determined, $\rho_0(\mathbf{r})$ is, say, the neutron density as a function of the coordinates, and $K(\mathbf{r},\mathbf{r}')$ is the kernel (assumed to be symmetric in $\mathbf{r}$ and $\mathbf{r}'$); the integration is over the medium in question. It can then be shown that the functional:

$$\int d\mathbf{r}\rho(\mathbf{r})\left\{\rho(\mathbf{r}) - c\int d\mathbf{r}'\rho(\mathbf{r}')K(\mathbf{r},\mathbf{r}')\right\} \tag{2}$$

is an extremum[1] for the actual solution ρ_0 of Eq. (1); furthermore, it is evident that:

$$c = \frac{\int d\mathbf{r}\rho_0^2(\mathbf{r})}{\int d\mathbf{r}\rho_0(\mathbf{r})\int d\mathbf{r}'\rho_0(\mathbf{r}')K(\mathbf{r},\mathbf{r}')}. \tag{3}$$

Since the first variation of (2) vanishes, we may expect that the value predicted for the eigenvalue c by the insertion of some trial function for the unknown ρ_0 will be closer to the correct value of c than the trial function is to ρ_0. It can be proved that the approximate eigenvalue is always larger than the correct eigenvalue.

The extension of the variational method to inhomogeneous integral equations is possible and leads to a simple procedure for obtaining fairly accurate solutions for a certain class of neutron diffusion problems. We shall first present briefly the method in general terms and then discuss specific applications. Suppose the inhomogeneous integral equation is:[2]

$$q_0(x) = \int dx'K(x,x')q_0(x') + f(x), \tag{4}$$

where $q_0(x)$ is the unknown function, $K(x,x')$ is a positive symmetric kernel, and $f(x)$ is bounded so that $\int|f(x)|dx$ exists. Suppose, further, that we can express $[\int dxq_0(x)]$ in terms of $q_0(\infty)$, say

$$q_0(\infty) = k_1\int dxq_0(x)f(x) + k_2$$

* This work was done in the Montreal Laboratory of the National Research Council of Canada during the Spring of 1944 and is now declassified.

[1] This can be seen easily by writing the trial function $\rho(\mathbf{r})$ as $[\rho_0(\mathbf{r})+\epsilon(\mathbf{r})]$, ($\epsilon(\mathbf{r})$ is the correction function), and showing that the term of first order in ϵ vanishes identically.

[2] The method is presented for a function of one variable; the generalization to more than one variable, as in the case of the homogeneous integral equations, is also possible.

with k_1, k_2 known constants. Let us define the functional:

$$\frac{\int dx q(x)\left\{q(x)-\int dx' K(x,x')q(x')\right\}}{\left[\int dx q(x)f(x)\right]^2}. \tag{5}$$

The reasoning then proceeds along the same lines as in the homogeneous case: (5) is an extremum[3] for the actual solution, $q_0(x)$, of Eq. (4), and can be equated to $[k_1/(q_0(\infty)-k_2)]$. For an arbitrary trial function, $q(x)$, the equation determining $q_0(\infty)$ is:

$$\frac{\int dx q(x)\left\{q(x)-\int dx' K(x,x')q(x')\right\}}{\left[\int dx q(x)f(x)\right]^2}=\frac{k_1}{q_0(\infty)-k_2}. \tag{6}$$

We shall now show how several neutron diffusion problems requiring a knowledge of the asymptotic neutron density can be written in the form (4). The application of (6) will follow immediately.

We consider first the well-known Milne problem, i.e., the case of a semi-infinite, isotropically scattering (but non-capturing) medium bounded by vacuum which sustains a constant current from infinity. A rigorous expression for the asymptotic neutron density is known for this problem;[4] this is not the case for the other two problems discussed below where at present only approximate solutions are available. The application of the variational method to the Milne problem will disclose the manner in which values of the asymptotic neutron density are arrived at and give an indication of the power of the method.

The integral equation for the neutron density, $\rho_0(x)$ (the boundary is taken at $x=0$ and the scattering mean free path is taken as the unit of length), in the Milne problem is:

$$\rho_0(x)=\tfrac{1}{2}\int_0^\infty dx' E_1(|x-x'|)\rho_0(x'), \tag{7}$$

where $E_1(x)$ is the exponential integral function of order one.[5] Since the asymptotic neutron density is linear, we write $\rho_0(x)=x+q_0(x)$; for large x, $q_0(x)$ approaches a constant which we denote by $q_0(\infty)$. Substitution for $\rho_0(x)$ into (7) yields:

$$q_0(x)=\tfrac{1}{2}\int_0^\infty dx' E_1(|x-x'|)q_0(x')+\frac{E_3(x)}{2}, \tag{8}$$

where $E_3(x)$ is the exponential integral function of order three. Equation (8) is of the form (4) but we still require an expression for $\int_0^\infty dx q_0(x)E_3(x)$ in terms of $q_0(\infty)$ before applying (6).

To obtain such a relation,[6] we make use of a theorem proved by Davison,[7] namely that if we have

---

[3] Dr. B. Davison, in the following article, discusses the conditions under which (5) is a minimum, not only an extremum, for the actual solution.

[4] E. Hopf, *Mathematical Problems of Radiative Equilibrium*, Cambridge Tracts No. 31 (1934).

[5] The exponential integral function of order n is: $E_n(x)=\int_1^\infty[e^{-xt}]dt/t^n$.

[6] In the original report, the relation between $q_0(\infty)$ and $\int_0^\infty dx q_0(x)f(x)$ in all three problems considered, was derived by another method. It is more elegant to make use of Davison's theorem.

[7] The proof of this theorem is simply as follows: $Q(r)$ in (9) can be regarded as r times the neutron density in an infinite (isotropically) scattering medium due to a spherically symmetrical system of continuously distributed sources, the source density at distance r from the origin being $F(r)/r$. The total output of these sources will be $\int dr(F(r)/r)=4\pi\int_0^\infty r dr F(r)$ and $Q(\infty)$ is equal to $3/4\pi$ times the output of all sources; this leads to (10).

an integral equation of the form:

$$Q(r)=\tfrac{1}{2}\int_0^\infty dr'Q(r')[E_1(|r-r'|)-E_1(r+r')]+F(r) \tag{9}$$

and $Q(r)$ is its solution bounded at infinity, then

$$Q(\infty)=3\int_0^\infty rdrF(r). \tag{10}$$

A comparison of Eqs. (8) and (9) yields:

$$q_0(\infty)=\tfrac{3}{2}\int_0^\infty xdxE_3(x)+\tfrac{3}{2}\int_0^\infty xdx\int_0^\infty dx'q_0(x')E_1(x+x'). \tag{11}$$

Interchanging the order of integration in the second term on the right-hand-side (r.h.s.) of (11) and evaluating the resulting integral and the first term on the r.h.s., we get:

$$q_0(\infty)=\tfrac{3}{2}\int_0^\infty dxq_0(x)E_3(x)+\tfrac{3}{8}. \tag{12}$$

Equation (12) is of the desired form and Eq. (6) becomes:

$$\frac{\int_0^\infty dxq(x)\left\{q(x)-\tfrac{1}{2}\int_0^\infty dx'q(x')E_1(|x-x'|)\right\}}{\tfrac{1}{2}\left[\int_0^\infty dxq(x)E_3(x)\right]^2}=\frac{1}{\tfrac{3}{8}q_0(\infty)-\tfrac{1}{4}}. \tag{13}$$

Equation (13) is the final result, and from it we may calculate $q_0(\infty)$ for different choices of the trial function $q(x)$. The procedure now is to assume a function for $q(x)$, and choose the parameters in this function so that the left-hand side (l.h.s.) of (10) is an extremum; the resulting function is then used to calculate $q_0(\infty)$. In this process the trial function $q(x)$ is determined except for a constant factor; this factor may be taken so that $q(x)$ is asymptotic to the value obtained for $q_0(\infty)$.

As has already been remarked, the whole point of the variational method is that the choice of $q(x)$ may be quite rough and yet give a fairly accurate value for $q_0(\infty)$. Thus, in the present instance, even if we assume $q(x)=$constant (for which no variation is necessary in the left-hand side of (13)), we get:

$$q_0(\infty)=\tfrac{3}{8}\left\{1+\frac{\left[\int_0^\infty dxE_3(x)\right]^2}{\int_0^\infty dx\left[1-\tfrac{1}{2}\int_0^\infty dx'E_1(|x-x'|)\right]}\right\} \tag{14}$$

$$=17/24=0.7083,$$

which is only 0.3 percent less than the correct value 0.7104.[8] The result can be considerably improved by starting from a more elaborate trial function and actually using the extremum property of the functional. In fact, LeCaine has shown that by choosing a trial function which simulates more closely the the actual $q(x)$, namely:

$$q(x)=\text{const.}[1-AE_2(x)+BE_3(x)]$$

[8] Cf. E. Hopf, reference 4; the value 0.7083 should be less than the true value (which it is!) since the l.h.s. of (13) should give a minimum for the correct solution of the integral equation (8) (cf. reference 3).

and extremizing the left-hand side of (13) with respect to A and B, one arrives at a value of $q_0(\infty)$ which agrees with the correct value to six decimal places (0.7104457 as compared to the correct value 0.7104461).

The second application of the variational method considered here is an estimate of the asymptotic neutron density in an infinite, isotropically scattering (but non-capturing) medium supporting a constant flux from infinity and surrounding a perfectly absorbing spherical core. This problem was treated by the spherical harmonic method,[9] and we shall compare the results of the variational method with those results. The integral equation for the neutron density in the medium surrounding the black core can easily be derived; we find:

$$\rho_0(r)=\tfrac{1}{2}\int_a^\infty dr'\rho_0(r')\{E_1(|r-r'|)-E_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})\}, \tag{15}$$

where $\rho_0(r)$ is r times the neutron density and a is the radius of the black core.

We are interested in a solution of (15) which is asymptotically linear, and we write this solution as

$$\rho_0(r)=r-Q(r). \tag{16}$$

The existence of this solution can be established in the same way as the existence of the solution of Eq. (7). On substituting (16) into (15), we find the integral equation for $Q(r)$, namely:

$$Q(r)=\int_a^\infty dr'Q(r')K(r,r')+\epsilon_3(r) \tag{17}$$

where

$$K(r,r')=\tfrac{1}{2}\{E_1(|r-r'|)-E_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})\}, \tag{17a}$$

$$\epsilon_3(r)=\tfrac{1}{2}[aE_2(r-a)+E_3((r^2-a^2)^{\frac{1}{2}})-E_3(r-a)]. \tag{17b}$$

The kernel $K(r,r')$ is obviously symmetric; it is also positive, as can be seen by noting that $|r-r'|<[(r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}}]$ and that $E_1(x)$ is a monotonically decreasing function of x. Moreover, the function $\epsilon_3(r)$ is bounded and the integral $\int_a^\infty|\epsilon_3(r)|dr$ exists. Hence, the variational method can be employed to estimate $Q(\infty)$ provided $Q(\infty)$ can be expressed in terms of $\int_a^\infty drQ(r)\epsilon_3(r)$.

A relation between $Q(\infty)$ and $\int_a^\infty drQ(r)\epsilon_3(r)$ can be found by again making use of Davison's theorem (cf. Eqs. (9) and (10)). For this purpose, we must define $Q(r)$ and $\epsilon_3(r)$ in the interval $0\leqslant r<a$; we choose:[10]

$$Q(r)=\tfrac{1}{2}\int_a^\infty dr'Q(r')\{E_1(|r-r'|)-E_1(r+r')\}, \quad (0\leqslant r<a), \tag{18a}$$

$$\epsilon_3(r)=0, \qquad (0\leqslant r<a). \tag{18b}$$

Further, we write:

$$\Lambda Q(r)=\tfrac{1}{2}\int_0^\infty dr'Q(r')\{E_1(|r-r'|)-E_1(r+r')\}, \tag{19a}$$

$$\Delta_1Q(r)=\tfrac{1}{2}\int_0^a dr'Q(r')\{E_1(|r-r'|)-E_1(r+r')\}, \tag{19b}$$

$$\Delta_2Q(r)=\tfrac{1}{2}\int_a^\infty dr'Q(r')\{E_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})-E_1(r+r')\}, \quad (r\geqslant a), \tag{19c}$$

$$=0, \quad (0\leqslant r<a). \tag{19d}$$

[9] R. E. Marshak, Phys. Rev. 71, 688 (1947).

[10] Cf. Davison's declassified Montreal report MT-232, "Influence of an Air Gap Surrounding a Small Black Sphere Upon the Linear Extrapolation Length of the Neutron Density in the Surrounding Medium."

In view of the above definitions, Eq. (17) becomes:

$$Q(r)=\Lambda Q(r)-\Delta_1 Q(r)-\Delta_2 Q(r)+\epsilon_3(r). \tag{20}$$

Equation (20) is also satisfied for $0\leqslant r<a$. If we now use Eq. (9), we get:

$$Q(\infty)=3\left\{\int_a^\infty rdr\epsilon_3(r)-\int_0^\infty rdr[\Delta_1 Q(r)]-\int_a^\infty rdr\Delta_2 Q(r)\right\}. \tag{21}$$

The evaluation of the first term on the r.h.s. of (21) is straightforward; the result is $3a^2/4$. The second term on the r.h.s. can be evaluated by interchanging the order of integration to give $-3\int_0^a rdrQ(r)$. However, $Q(r)$ for $0\leqslant r<a$ is defined by (18a); inserting (18a) and interchanging the order of integration a second time yields:

$$-\tfrac{3}{2}\int_a^\infty dr'Q(r')\int_0^a rdr\{E_1(|r-r'|)-E_1(r+r')\}.$$

The third term on the r.h.s. becomes, on interchanging the order of integration:

$$-\tfrac{3}{2}\int_a^\infty dr'Q(r')\int_a^\infty rdr\{E_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})-E_1(r+r')\}.$$

It is easy to show that:

$$\tfrac{1}{2}\left\{\int_0^a rdrE_1(|r-r'|)+\int_a^\infty rdrE_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})-\int_0^\infty rdrE_1(r+r')\right\}=\epsilon_3(r'). \tag{22}$$

Hence, Eq. (21) leads to the desired relation:

$$Q(\infty)=\frac{3a^2}{4}-3\int_a^\infty drQ(r)\epsilon_3(r). \tag{23}$$

Substituting (23) into (6) yields an equation corresponding to (13), namely:

$$\frac{\int_a^\infty drQ(r)\left\{Q(r)-\frac{1}{2}\int_a^\infty dr'Q(r')K(r,r')\right\}}{\left[\int_a^\infty drQ(r)\epsilon_3(r)\right]^2}=\frac{1}{\frac{a^2}{4}-\frac{Q(\infty)}{3}}. \tag{24}$$

Again, as in the plane Milne case, we choose the simplest trial function, i.e. $Q(r)\equiv$const.; we get

$$Q(\infty)=\frac{3a^2}{4}-\frac{3\left[\frac{a}{2}-\frac{1}{3}+\int_a^\infty drE_3((r^2-a^2)^{\frac{1}{2}})\right]^2}{\left[1+2\int_a^\infty dr\int_a^\infty dr'E_1((r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}})\right]}. \tag{25}$$

Values of $[a-Q(\infty)]$—i.e., the "extrapolated endpoint"—as a function of a computed on the basis of (25) are given in column 2 of Table I, while the values of the extrapolated endpoint predicted by the P_6-approximation of spherical harmonic method[9] are given in column 3. It is seen how powerful the variational method is for the determination of the asymptotic neutron density. Of course, a

slightly more accurate choice of trial function, which takes into account the exponential-type approach of $Q(r)$ to $Q(\infty)$, would lead to considerably more accurate values of the extrapolated endpoint. Table I further indicates that the actual solution of (17) maximizes the extrapolated endpoint as is to be expected from Davison's paper.[3]

TABLE I. Extrapolation endpoint as function of a.

a	Variational method	P_3-approximation
0.5	0.347	0.334
1.0	.506	.494
2.0	.626	.620
5.0	.688	.690
∞	.708	.709

Finally, we have applied the variational method developed above to the case of a black sphere (of radius a) surrounded by an "air gap" (of outer radius b) which in turn is surrounded by an isotropically scattering (but non-capturing) medium supporting a constant flux from infinity. The integral equation for r times the neutron density turns out to be:

$$\rho_0(r)=\tfrac{1}{2}\int_b^\infty dr'\,\rho_0(r')\bar{K}(r,r') \tag{26}$$

where:

$$\bar{K}(r,r')=\left\{E_1(|r-r'|)-E_1((r^2-b^2)^{\frac{1}{2}}+(r'^2-b^2)^{\frac{1}{2}})+\int_\alpha^\beta \frac{dR}{R}\exp\left[-R+\frac{1}{R}((R^2+r^2-r'^2)^2-4R^2(r^2-b^2))^{\frac{1}{2}}\right]\right\}.$$

with

$$\alpha=(r^2-b^2)^{\frac{1}{2}}+(r'^2-b^2)^{\frac{1}{2}},\quad \beta=(r^2-a^2)^{\frac{1}{2}}+(r'^2-a^2)^{\frac{1}{2}}.$$

As in the two preceding cases, since capture is absent in the outside medium, the asymptotic solution is linear. Following the procedure for the black sphere without gap, we obtain the integral equation for $Q(r)$, $(\rho_0(r)=r-Q(r))$, namely:

$$Q(r)=\tfrac{1}{2}\int_b^\infty dr'Q(r')\bar{K}(r,r')+\bar{t}_3(r), \tag{27}$$

where:

$$\bar{t}_3(r)=\tfrac{1}{2}[bE_2(r-b)-E_3(r-b)+E_3((r^2-a^2)^{\frac{1}{2}}-(b^2-a^2)^{\frac{1}{2}})-(b^2-a^2)^{\frac{1}{2}}E_2((r^2-a^2)^{\frac{1}{2}}-(b^2-a^2)^{\frac{1}{2}})]. \tag{27a}$$

We find for $Q(\infty)$ (cf. Eq. (23)):

$$Q(\infty)=\frac{3a^2}{4}-3\int_b^\infty drQ(r)\bar{t}_3(r). \tag{28}$$

Substitution of (28) into Eq. (6) then permits an immediate application of the variational method to estimate $Q(\infty)$. No numerical results have been obtained for this case.

It is a pleasure to thank Professor J. Schwinger for a valuable conversation on variational methods. Dr. Davison's cooperation is also appreciated.

A Remark on the Variational Method

B. DAVISON
Chalk River Laboratory, Chalk River, Ontario
(Received January 24, 1947)

The nature of the extremum of the functional

$$\int q(x)\left[q(x)-\lambda\int q(x)K(x,y)dy\right]dx\Big/\left[\int q(x)f_0(x)dx\right]^2,$$

associated with the inhomogeneous integral equation,

$$q_0(x)=\lambda\int q_0(y)K(x,y)dy+f_0(x),$$

is investigated. Two sets of sufficient conditions for the extremum reached for $q(x)=q_0(x)$ to be a minimum (and neither a maximum, nor a "saddle point") are determined.

1. INTRODUCTION

DR. R. E. MARSHAK in the article on "The Variational Method for Asymptotic Neutron Densities," appearing in the same issue of this journal, solves certain inhomogeneous integral equations by use of the following theorem:

Let $q_0(x)$ be the (bounded) solution of the integral equation

$$q_0(x)-\lambda\int_0^\infty K(x,y)q_0(y)dy-f_0(x)=0, \quad (1)$$

in which the parameter λ is positive and its value is known beforehand, the kernel $K(x,y)$ is positive and symmetrical and the free term $f_0(x)$ is bounded, and the integral $\int_0^\infty |f_0(x)|dx$ exists.

Then the functional

$$I=\frac{\int_0^\infty q(x)\left[q(x)-\lambda\int_0^\infty K(x,y)q(y)dy\right]dx}{\left[\int_0^\infty q(x)f_0(x)dx\right]^2}, \quad (2)$$

in which $q(x)$ is any bounded real well-behaved function, reaches its minimum for $q(x)=q_0(x)$, provided either

$$\lambda\int_0^\infty K(x,y)dy\leqslant 1 \quad (3)$$

for any fixed x, or

$$\lambda<\lambda_0 \quad (4)$$

in which λ_0 is the limit, for $b\to\infty$, of the lowest eigenvalue of the equation

$$\varphi(x)=\lambda'\int_0^b K(x,y)\varphi(y)dy, \quad (5)$$

and provided $K(x,y)$ satisfies certain general qualitative conditions, which will be formulated later.

The purpose of the present note is to prove this theorem.

(Notice that the conditions (3) and (4) are not equivalent, but represent two different types of cases in which one can use the functional (2): There are cases when $\lambda=\lambda_0$ but (3) is satisfied, and there are cases when (4) is satisfied but (3) is violated.)

We shall split the proof of the theorem into two parts. In the first part we prove the lemma:

The functional (2) reaches its minimum for $q(x)=q_0(x)$, if λ and $K(x,y)$ are such that

$$\int_0^\infty \bar{q}(x)\left[\bar{q}(x)-\lambda\int_0^\infty K(x,y)\bar{q}(y)dy\right]dx \quad (6)$$

is positive for any bounded, not identically vanishing, real, well-behaved $\bar{q}(x)$, and that

$$\int_0^\infty q_1(x)f_0(x)dx=\int_0^\infty q_0(x)f_1(x)dx \quad (7)$$

in which $f_1(x)$ is any function satisfying the same assumptions as $f_0(x)$ and $q_1(x)$ is the solution of the Eq. (1) with $f_0(x)$ replaced by $f_1(x)$.

In the second part, under some general qualitative assumptions with respect to $K(x,y)$ we prove that (6) is positive for any bounded, not

identically vanishing, real, well-behaved $\tilde{q}(x)$, if either (3) or (4) is satisfied.

2. PROOF OF THE LEMMA

Let $q(x)$ be some trial function which we substitute into the functional (2). Present it in the form

$$q(x)=q_0(x)+\epsilon q_1(x). \tag{8}$$

Let $\epsilon f_1(x)$ be the value of the left-hand side of (1) when we replace there $q_0(x)$ by $q(x)$ (i.e. $q_1(x)$ is the solution of (1) with $f_0(x)$ replaced by $f_1(x)$), and substitute (8) into (2). Then, after a slight rearrangement, and on using (7), we get

$$I=\frac{1}{\int_0^\infty q_0(x)f_0(x)dx}\left\{1+\left[\frac{\epsilon}{\int_0^\infty q(x)f_0(x)dx}\right]^2\Delta\right\}, \tag{9}$$

in which

$$\Delta=\begin{vmatrix}\int_0^\infty f_0(x)q_0(x)dx & \int_0^\infty f_0(x)q_1(x)dx\\ \int_0^\infty f_1(x)q_0(x)dx & \int_0^\infty f_1(x)q_1(x)dx\end{vmatrix}, \tag{10}$$

or, which is the same as,

$$\Delta=\int_0^\infty f_0(x)q_0(x)dx \cdot\int_0^\infty[f_1(x)-\mu f_0(x)][q_1(x)-\mu q_0(x)]dx, \tag{11}$$

in which μ is a constant determined from the condition

$$\int_0^\infty q_0(x)[f_1(x)-\mu f_0(x)]dx=0.$$

Put $q_1(x)-\mu q_0(x)=\tilde{q}(x)$, and notice that thereby

$$f_1(x)-\mu f_0(x)=\tilde{q}(x)-\lambda\int_0^\infty K(x,y)\tilde{q}(y)dy.$$

Thus if (6) is necessarily positive then Δ is positive, and the expression (9) reaches its minimum for $\epsilon=0$, i.e., the functional (2) reaches its minimum for $q(x)=q_0(x)$, which we wanted to prove.

Notice that if λ and $K(x,y)$ are such that (6) can have either sign, depending on the actual form of $\tilde{q}(x)$, then, assuming $\int_0^\infty f_0(x)q_0(x)dx$ to be positive, Δ can have either sign depending on what trial function we used, and (9) will have neither a minimum nor a maximum, but a "saddle-point" for $\epsilon=0$, i.e., the functional (2) will have neither a minimum nor a maximum for $q(x)=q_0(x)$, but a "saddle-point."

3. CASE WHEN $\lambda<\lambda_0$

Turn now to the determination of conditions under which (6) is positive for any not identically vanishing, real, well-behaved function $\tilde{q}(x)$, and let us deal first with the case $\lambda<\lambda_0$.

Let us examine first the case of an equation with a finite range of integration, i.e., the case when $K(x,y)$ and $f(x)$ vanish for $x>b$, so that the homogeneous equation corresponding to our inhomogeneous has the form (5). Let λ_0', λ_1', etc., be the complete set of eigenvalues for the Eq. (5) and $\varphi_0(x)$, $\varphi_1(x)$, etc., be the corresponding eigenfunctions. It is known that since $K(x,y)$ is symmetrical, all λ_m' should be real and the $\varphi_m(x)$ should be mutually orthogonal.[1] We shall assume further that $K(x,y)$ is such that all λ_m' are positive and $\varphi_m(x)$ form a complete set of orthogonal functions. Let λ_0' be smallest of all λ_m' (so that $\lim_{b\to\infty}\lambda_0'=\lambda_0$). Once we have assumed that $\varphi_m(x)$ form a complete set of orthogonal functions, we can write

$$\tilde{q}(x)=\sum_{n=0}^{\infty}a_n\varphi_n(x) \tag{12}$$

and hence[2]

$$\tilde{q}(x)-\lambda\int_0^b K(x,y)\tilde{q}(y)dy=\sum_{n=0}^{\infty}a_n\frac{\lambda_n'-\lambda}{\lambda_n'}\varphi_n(x),$$

so that

$$\int_0^b\tilde{q}(x)\left[\tilde{q}(x)-\lambda\int_0^b K(x,y)\tilde{q}(y)dy\right]dx=\sum_{n=0}^{\infty}a_n^2\frac{\lambda_n'-\lambda}{\lambda_n'}\int_0^b\varphi_n^2(x)dx. \tag{13}$$

[1] See e.g., Whittaker and Watson, *Modern Analysis* (Cambridge University Press, Cambridge, England, 1927), pp. 225–227.

[2] See reference 1, p. 228.

If $\lambda<\lambda_0$, then for sufficiently large values of b we shall also have $\lambda<\lambda_0'$, hence all the terms on the right-hand side of (13) will be positive. Then the left-hand side of (13) should be positive too, and passing to the limit with $b\to\infty$ we see that (6) is positive too.

If $\lambda>\lambda_0$, one can see from the same argument that (13) and consequently (6) will be negative if $\bar{q}(x)$ is such that the term in (12) corresponding to the lowest eigenvalue [or the lower eigenvalues] strongly predominates over the terms corresponding to high eigenvalues, and positive in the opposite case, so that for $\lambda>\lambda_0$, the quantity (6) can have either sign depending upon the form of $\bar{q}(x)$, and hence the functional (2) will have merely a saddle-point for $q(x)=q_0(x)$. The case of $\lambda=\lambda_0$ should be considered separately, and this will be done in the next section.

Insofar as the assumption (7) used in proving the lemma is concerned, it may be pointed out that if $\lambda<\lambda_0$ and if $K(x,y)$ is such that, in the integral

$$\int_0^\infty\cdots\int_0^\infty f_0(x)K(x,s_1)K(s_1,s_2)\cdots \times K(s_n,y)f_1(y)dxds_1\cdots ds_ndy, \quad (14)$$

we can change the order of integration in any arbitrary manner, then (7) is automatically satisfied. For if $\lambda<\lambda_0$, then $q_0(x)$ and $q_1(x)$ can be determined by means of Neumann's procedure, once $f_0(x)$ and $f_1(x)$ are known, i.e.,

$$q_0(x)=f_0(x)+\lambda\int_0^\infty K(x,y)f_0(y)dy +\lambda^2\int_0^\infty\int_0^\infty K(x,s)K(s,y)f_0(y)dyds+\text{etc.}, \quad (15)$$

and a similar expression for $q_1(x)$, the series appearing in both these expansions being absolutely and uniformly convergent. And then from the possibility of interchanging the order of integration in (14), (7) follows at once.

4. CASE WHEN $\lambda=\lambda_0$

In this case we can prove that (6) is positive for any bounded, not identically vanishing, real, well-behaved $\bar{q}(x)$, and consequently (2) reaches its minimum for $q(x)=q_0(x)$, provided (3) is satisfied, and provided $\lambda K(x,y)$ satisfies the following additional assumptions:[3]

(i) The homogeneous equation corresponding to (1) cannot have any bounded solutions (so that the bounded solution of the inhomogeneous equation is necessarily unique)

(ii) $f(x)$ is bounded and the integral $\int_0^\infty|f(x)|dx$ exists

(iii) For any $f(x)$ satisfying (ii) the series of the form (15) converges for any particular value of x and its sum is uniformly bounded with respect to x; in other words, using Hopf's terminology,[4] the Eq. (1) possesses a unique bounded N-solution for any $f(x)$ satisfying (ii)

(iv) If $f_1(x)$ and $f_2(x)$ satisfy (ii), and $q_2(x)$ is the solution of (1) with $q_0(x)$ replaced by $q_2(x)$, then in the integrals

$$\int_0^\infty\cdots\int_0^\infty f_1(x)K(x,s)\cdots \times K(s_n,y)q_2(y)dyds_n\cdots dsdx, \quad (14')$$

we can interchange the order of integration in any arbitrary manner.

Notice that (7) follows at once from our equations and the assumption (iv). Notice also the following corollaries of assumptions (i) and (iii).

(A) If the free term is non-negative, then the corresponding bounded solution is also non-negative. (This follows since every term in the series of the form (15) is non-negative, consequently its sum is also non-negative. On the other hand this sum is obviously a solution, and according to (i) the solution is unique.)

(B) Put for brevity

$$\Lambda f(x)=\lambda\int_0^\infty K(x,y)f(y)dy; \qquad \Lambda\Lambda^{n-1}f(x)=\Lambda^n f(x). \quad (16)$$

Then

$$\lim_{n\to\infty}\Lambda^n f(x)=0 \quad (x \text{ fixed}). \quad (17)$$

---

[3] Since our purpose is merely to give a rigorous justification of the method of solution of certain problems given in Marshak's paper, we shall not attempt to prove our result under the most general assumptions possible, nor to make our assumptions mutually independent. But we shall assume such properties of $K(x,y)$ as it does possess in the examples treated by Marshak, and shall assume all the required properties independently, irrespective of the fact that some of them may be corollaries of the others.

[4] E. Hopf, *Mathematical Problems of Radiative Equilibrium* (Cambridge Tracts on Mathematics and Mathematical Physics, 1934).

(This is merely the necessary condition for the convergence of the series (15) with $f_0(x)$ replaced by $f(x)$.)

It should be mentioned that the only way in which we shall use the assumption (iii) is through its corollaries (A) and (B). It should also be mentioned with respect to (B) that we shall not use this corollary in full, but merely a somewhat weaker and clearer property, namely:

(B′) Either one of the consecutive terms in the series (15) with $f_0(x)$ replaced by $f(x)$ becomes non-negative (or non-positive), or for any particular value of x the consecutive terms of this series tend to zero (i.e., we shall need to use (17) only for the cases when none of the functions $\Lambda^n f(x)$ conserves its sign throughout the range of x-values).

Having clarified our assumptions, turn now to the proof that (6) is necessarily positive. Put

$$\tilde{f}(x)=\tilde{q}(x)-\lambda\int_0 K(x,y)\tilde{q}(y)dy, \tag{18}$$

and

$$N(\tilde{f})=\int_0^\infty \tilde{f}(x)\tilde{q}(x)dx, \tag{19}$$

and notice that $\tilde{f}(x)=f_1(x)-\mu f_0(x)$ satisfies the assumption (ii) and hence, according to (3), so do all $\Lambda^n\tilde{f}(x)$. Using the assumption (iv), and noticing that $\tilde{q}(x)$ is the bounded solution of (1) with $f_0(x)$ replaced by $\tilde{f}(x)$ and $\Lambda\tilde{q}(x)$ is the bounded solution of (1) with $f_0(x)$ replaced by $\Lambda\tilde{f}(x)$ we have

$$N(\tilde{f})=\int_0^\infty \tilde{f}^2(x)dx+\lambda\int_0^\infty\int_0^\infty K(x,y)\times\tilde{f}(x)\tilde{f}(y)dxdy+N(\Lambda\tilde{f}). \tag{20}$$

The first term on the right-hand side in (20) is obviously positive. Let us show that the absolute value of the second term is less than the first term. Use for this purpose Schwartz's inequality, which in two dimensions has the form

$$\iint F_1^2(x,y)dxdy\cdot\iint F_2^2(x,y)dxdy>\left[\iint F_1(x,y)F_2(x,y)dxdy\right]^2.$$

$F_1(x,y)$ and $F_2(x,y)$ are real. Taking here

$$F_1(x,y)=\tilde{f}(x)\lambda K(x,y),\quad F_2(x,y)=\tilde{f}(y)\lambda K(x,y),$$

interchanging in the integral $\int\int F_2^2(x,y)dxdy$ the variables x and y, and extracting the square root, we get

$$\left|\lambda\int_0^\infty\int_0^\infty K(x,y)\tilde{f}(x)\tilde{f}(y)dxdy\right|<\lambda\int_0^\infty\int_0^\infty K(x,y)\tilde{f}^2(x)dxdy. \tag{21}$$

On the other hand multiplying (3) by $\tilde{f}^2(x)$ and integrating over all x, we see that the right-hand side of (21) is less than $\int_0^\infty\tilde{f}^2(x)dx$. Then it follows from (20) that

$$N(\tilde{f})>N(\Lambda\tilde{f}), \tag{22}$$

and similarly,

$$N(\Lambda\tilde{f})>N(\Lambda^2\tilde{f}),\quad \text{etc.}$$

If one of the functions $\tilde{f}(x)$, $\Lambda\tilde{f}(x)$, $\Lambda^2\tilde{f}(x)$, etc., $\Lambda^m\tilde{f}(x)$, say, conserves its sign then $N(\Lambda^m\tilde{f})$ is positive, and, according to (22), so is $N(f)$. If none of the functions $\Lambda^n\tilde{f}(x)$ conserves its sign, then, according to (B′), $\Lambda^n\tilde{f}(x)$ will tend to zero with increasing n, and $N(\Lambda^n\tilde{f}(x))$ will also tend to zero with increasing n, and the formulae (22) then show again that $N(\tilde{f})$ is positive, i.e., the expression (6) is positive and the functional (2) reaches its minimum for $q(x)=q_0(x)$.

The True Height of an Ionospheric Layer

J. A. PIERCE
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received November 4, 1946)

A simple method has been developed for the analysis of ionospheric sweep-frequency records in terms of the scale height and true height of maximum of the layer of Chapman form which most closely fits the observed data. It is shown that the height of maximum determined by the method of Appleton, Booker, and Seaton is the height of a parabolic layer which is not coincident with the parabola that most closely fits the Chapman distribution. The present method includes a correction for retardation in one or more layers of similar form which lie below the reflecting layer.

INTRODUCTION

BOOKER and Seaton[1] have published a discussion of a method for analyzing virtual height-frequency records in terms of the assumption of a parabolic distribution of densities, in order to determine the true height of the maximum and the relative thickness of the distribution curve. It is the purpose of the present paper to develop a similar treatment for ionospheric data of the same type, which is based upon the assumption of a Chapman distribution.

The assumed variation of the density of ionization with height in an ionospheric layer is shown by the solid curve of Fig. 1. The relative density is given by the equation

$$N/N_0=[\exp(a+1-e^a)]^{\frac{1}{2}} \qquad (1)$$

where N=ion density at a height h, N_0=maximum ion density, e=base of the natural logarithms, $a=(h_M-h)/H$, h_M=height at which $N=N_0$, and H=scale height. The constants used for the curve shown are $h_M=300$ km and $H=50$ km. Equation (1) is given by G. W. Pierce[2] and defines the Chapman distribution in ratio form. Its derivation is based on the concept of monochromatic ionizing radiation falling upon an atmosphere of constant temperature with a density that decreases exponentially with increasing height. The equation actually defines the rate of production of ions under these conditions and H is the height of the homogeneous atmosphere. As diffusion, among other factors, is not considered, we may not leap to the con-

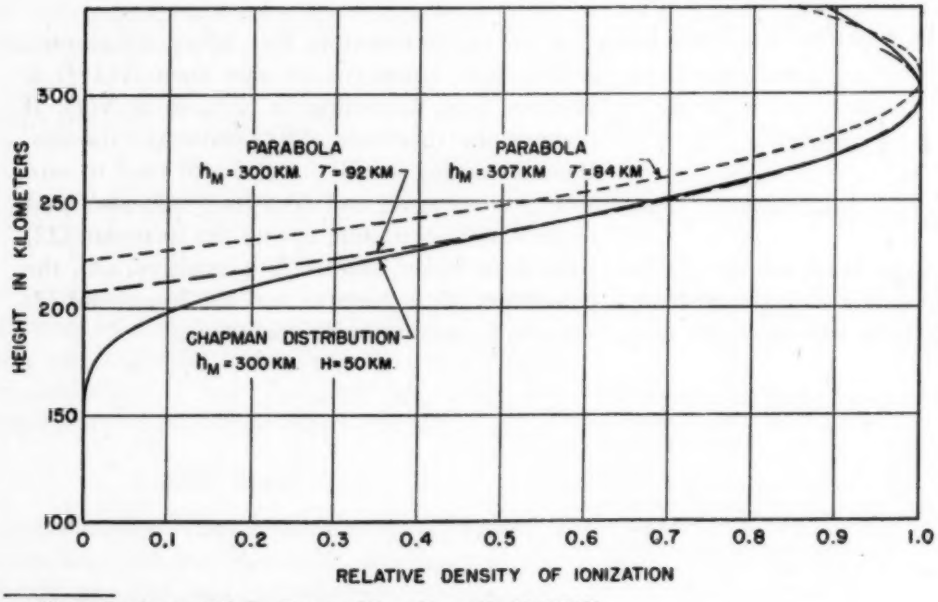


FIG. 1. Comparison between Chapman and parabolic distributions of ionization.

[1] H. G. Booker and S. L. Seaton, Phys. Rev. 57, 87 (1940).
[2] G. W. Pierce, Cruft Laboratory Lectures.

clusion that the total density of ionization at any moment is defined by this equation, but it seems reasonable to suppose that the sustained levels of ionization will be represented by a curve of this same general form. It will be assumed, for the time being, that the sustained density has exactly this form. It does not follow, however, that the scale height, H, for the density distribution is exactly the height of the homogeneous atmosphere.

A parabolic distribution is defined by the equation,

$$\frac{N}{N_0}=1-\left(\frac{h_M-h}{\tau}\right)^2, \tag{2}$$

where τ is the half-thickness of the parabola (where $N=0$) and the other quantities have been defined under Eq. (1). It will be seen from Fig. 1 that the Chapman distribution may be fitted quite closely by a parabola, when N/N_0 is greater than about 0.4. In comparing the parabola with the Chapman curve, we find that if $\tau=1.82H$ the parabola lies wholly within the other curve and is approximately tangent to it at the point where $N/N_0=0.710$. If $\tau=2H$, the parabola is coincident with the Chapman distribution at the maximum and the curves cross at the point where $N/N_0=0.206$. The parabola shown is drawn for $\tau=1.84H$ and is considered to be about the best over-all fit.

Synthesis of a virtual height-frequency curve on the assumption of a parabolic distribution of density neglects the retardation due to passage of the rays through the lower part of the Chapman distribution, a part that must certainly exist in the ionosphere. This retardation will be small, but not negligible, for frequencies near the critical frequency, and will be of considerable importance at the lower frequencies where the rays are reflected at lower heights.

THE METHOD OF COMPUTATION

For the calculation of the virtual heights we shall make the assumptions that propagation is in a region of no magnetic field (or that transmission is for the ordinary ray in a direction perpendicular to the field) and that the group velocity of the wave is the product of the velocity of light and the index of refraction. As our computation is to be made for the case of normal incidence upon the ionized region, the virtual height will be the integral of $1/\mu$ from $h=0$ to the height where $\mu=0$. Unfortunately a solution of this equation has been found for only a few distributions of density of ionization which do not include that defined in Eq. (1). It is therefore necessary to adopt an approximate method. The most straightforward treatment appears to be the computation of the group velocity, or preferably its reciprocal, for a large number of laminations each of which is assumed to have uniform density throughout its thickness. This computation must be made for each lamination at each of a fairly large number of frequencies in order to define a virtual height-frequency curve, but fortunately the formula for $1/\mu$ may be put into a form suitable for a calculation of this sort so that a virtual height for a single frequency may be determined, for a case involving a hundred or more laminations, within twenty or thirty minutes after the density values for the various laminations have been found.

The most convenient form for the equation is,

$$\frac{1}{\mu}=\left[\frac{(f/f_0)^2}{(f/f_0)^2-N/N_0}\right]^{\frac{1}{2}}, \tag{3}$$

where f_0 is the critical frequency or the frequency for which μ is zero when $N=N_0$. After tabulating values of N/N_0 for the various laminations in a vertical column, the values of $1/\mu$ can easily be worked out with a comptometer and set down in a parallel column which is preferably headed by the value of $(f/f_0)^2$. Since the value of $1/\mu$ is defined by the ratio $(N/N_0)/(f/f_0)^2$, the computation may be shortened by preparing a table relating these two quantities. This table need not be carried beyond the point where $1/\mu$ is 2.5 or 3.0 as the labor involved in carrying the table further would become greater than that saved by its use.

Because $1/\mu$ goes to infinity at the point of reflection, special consideration must be given to the last few laminations. One method is to take thinner and thinner laminations as the point of reflection is approached, but this is quite laborious and only defers the question of how to treat the last layer of all. A rough method, which has no theoretical justification but which seems

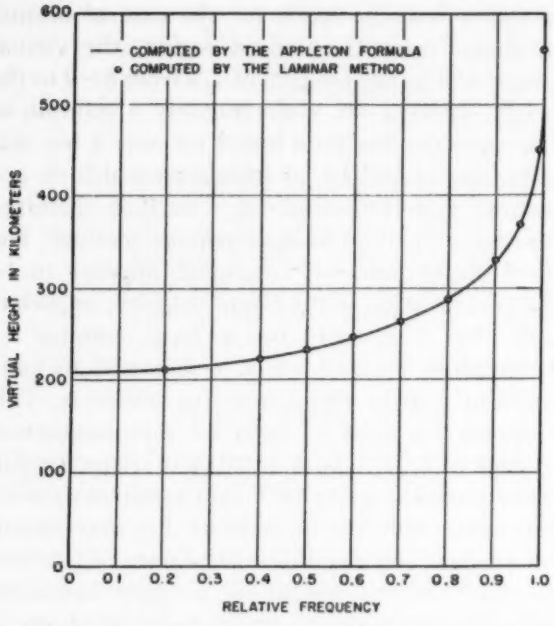


FIG. 2. Comparison between the laminar computation and an accurate method of calculating virtual heights. The data are for a parabolic distribution at a height of 300 km with a half-thickness of 92 km.

accurate enough for practical use, has been developed by a study of the parabolic distribution for which virtual heights may be calculated precisely. This method is to compute the value of N/N_0 for a point eight-tenths of the way from the top of the last complete lamination to the point of reflection (where $f^2/f_0^2=N/N_0$) and to multiply the corresponding value of $1/\mu$ by the thickness of this last fractional lamination. That this quantity, when added to the sum of the tabulated values of $1/\mu$ for the complete laminations, gives approximately the correct total virtual height is indicated by the fact that the treatment may alternatively be applied to the last one-and-a-fraction laminations instead of to the last fraction alone, with an alteration in the total virtual height which is generally not more than $0.02H$. The errors of the method are indicated in Fig. 2, where the solid curve is the relation between virtual height and relative frequency for a parabolic distribution of density ($h_M=300$ km, $\tau=92$ km) calculated from the equation given by Appleton,[3] and the circles are points computed as described above. It appears that the errors in these calculations, which were made for lamina of 1-km thickness, are small in comparison with the usual experimental error in measuring virtual heights.

METHODS OF ANALYSIS

For any law relating density of ionization to change in height it is necessary to calculate only one virtual height-frequency curve because the

TABLE I. Values of $(h'-h_M)/H$ as a function of f/f_0.

f/f_0	$\frac{h'-h_M}{H}$	f/f_0	$\frac{h'-h_M}{H}$	f/f_0	$\frac{h'-h_M}{H}$	f/f_0	$\frac{h'-h_M}{H}$	f/f_0	$\frac{h'-h_M}{H}$
0.00	$-\infty$	0.22	−1.92	0.44	−1.32	0.65	−0.71	0.85	0.25
0.01	−2.99	0.23	−1.89	0.45	−1.30	0.66	−0.67	0.86	0.32
0.02	−2.83	0.24	−1.86	0.46	−1.27	0.67	−0.63	0.87	0.40
0.03	−2.73	0.25	−1.84	0.47	−1.24	0.68	−0.60	0.88	0.48
0.04	−2.65	0.26	−1.81	0.48	−1.22	0.69	−0.56	0.882	½
0.05	−2.58	0.27	−1.78	0.49	−1.19	0.70	−0.52	0.89	0.58
0.06	−2.52	0.28	−1.75	0.50	−1.16	0.704	−½	0.90	0.68
0.07	−2.46	0.29	−1.73	0.51	−1.14	0.71	−0.48	0.91	0.79
0.08	−2.41	0.30	−1.70	0.52	−1.11	0.72	−0.44	0.92	0.91
0.09	−2.37	0.31	−1.67	0.53	−1.08	0.73	−0.40	0.927	1
0.10	−2.33	0.32	−1.65	0.54	−1.05	0.74	−0.35	0.93	1.04
0.11	−2.29	0.33	−1.62	0.55	−1.02	0.75	−0.31	0.94	1.20
0.12	−2.25	0.34	−1.59	0.558	−1	0.76	−0.26	0.95	1.39
0.13	−2.21	0.35	−1.57	0.56	−0.99	0.77	−0.21	0.955	1½
0.14	−2.17	0.36	−1.54	0.57	−0.96	0.78	−0.16	0.96	1.62
0.15	−2.14	0.37	−1.51	0.58	−0.93	0.79	−0.11	0.97	1.92
0.16	−2.10	0.38	−1.49	0.59	−0.90	0.80	−0.06	0.972	2
0.17	−2.07	0.39	−1.46	0.60	−0.87	0.81	0.00	0.98	2.34
0.18	−2.04	0.40	−1.43	0.61	−0.84	0.811	0	0.983	2½
0.19	−2.01	0.41	−1.41	0.62	−0.81	0.82	+0.05	0.99	3.04
0.20	−1.98	0.42	−1.38	0.63	−0.77	0.83	0.11	1.00	∞
0.21	−1.95	0.43	−1.35	0.64	−0.74	0.84	0.18		

[3] E. V. Appleton, Proc. Roy. Soc. A162, 451 (1937).

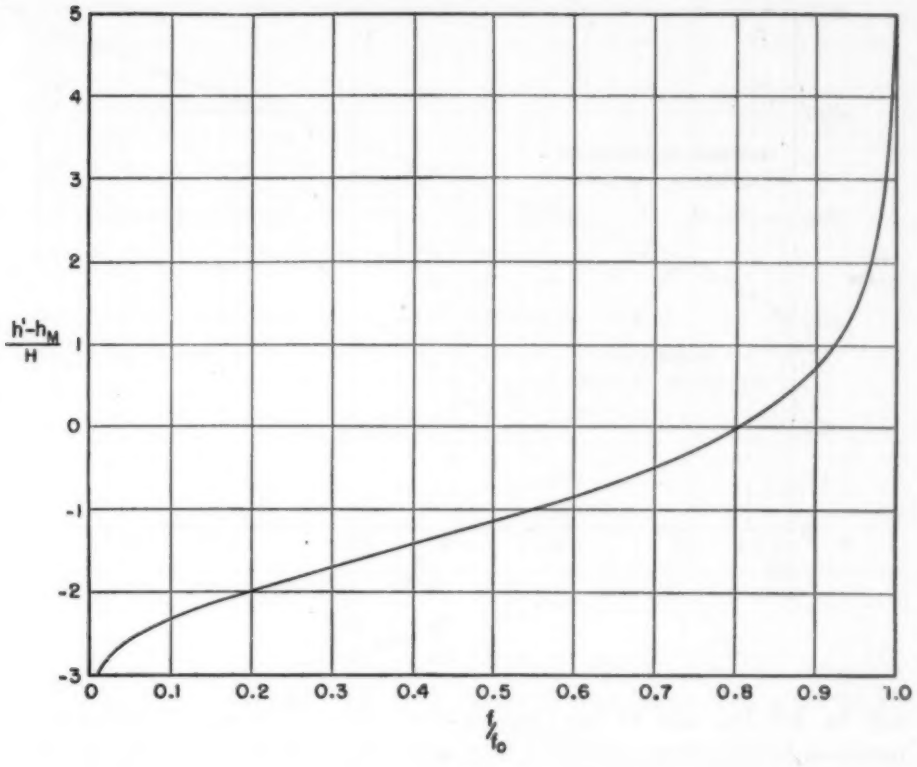


FIG. 3. The relation between virtual height and frequency for a Chapman distribution.

increase of virtual over true height of reflection is proportional to the scale height, as is the difference between true height of reflection and the height at which maximum density occurs. Thus the virtual height may be referred to the height of maximum density by the equation

$$h' = h_M + H\left[\frac{h'-h_M}{H}\right], \tag{4}$$

where h' is the virtual height of reflection and the other quantities have been defined above. The quantity in brackets is a function of f/f_0 which Booker and Seaton[1] have called "φ" when dealing with a parabolic region, but which we shall leave unnamed as it may assume any number of forms corresponding to various laws connecting the density of ionization with height. The relation between $(h'-h_M)/H$ and f/f_0 which has been calculated for the Chapman distribution is given in the graph of Fig. 3 and in Table I.

Appleton and Beynon[4] have noted that the virtual heights of reflection from a parabolic distribution fall in a straight line when plotted against

$$\frac{f}{2f_0}\log_e\frac{f_0+f}{f_0-f}.$$

The curve of Fig. 3 gives a line which is approximately straight, in the region where f/f_0 is greater than about 0.4, when plotted against this function of f and f_0. Because of the retardation in the lower part of the Chapman layer the straight part of the curve lies above the line for the equivalent parabola and has a slightly smaller slope, as shown in Fig. 4. This means that if the analysis for a parabolic distribution is applied to the virtual height-frequency record obtained from an ionospheric layer which actually has the Chapman form, it will result in values for the true height of maximum and the half-thickness which are respectively too large and too small. Thus if the virtual height-frequency curve for a Chapman layer with $h_M = 300$ km and $H = 50$ km is constructed from Eq. (4) and Fig. 3, and then the Booker and Seaton form of the parabolic treatment applied to the curve, the resulting values of h_M and τ

[4] E. V. Appleton and W. J. G. Beynon, Proc. Phys. Soc. 52, 518 (1940).

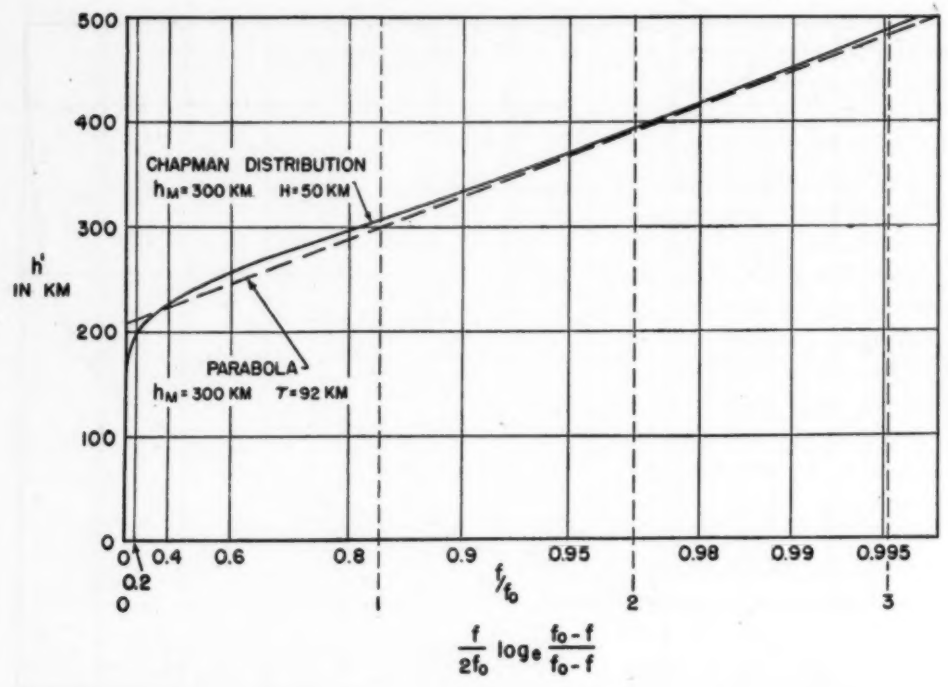


FIG. 4. Comparison of the virtual height-frequency curves for the distributions of ionization shown in Fig. 1.

will be 307 km and 84 km, respectively. The parabola having these constants is shown by the dotted curve of Fig. 1.

In practice the Booker and Seaton[1] method for the parabolic distribution may be converted into an analysis for the Chapman region by these rules:

1. Determine h_{Mp} and τ_p by the method given in reference 2. (Note: Subscripts have been used to differentiate between the values derived by the "parabolic" and "Chapman" treatments.)
2. Apply the following corrections:

a. $H=0.60\tau_p$,
b. $h_{Mc}=h_{Mp}-0.14H$,
c. $\tau_c=1.10\tau_p$.

TABLE II. Values of R/H as a function of f/f_0.

f/f_0	R/H	f/f_0	R/H	f/f_0	R/H	f/f_0	R/H
1.001	12.56	1.04	5.26	1.18	2.59	2.2	0.47
1.0015	11.74	1.05	4.85	1.19	2.51	2.4	0.39
1.002	11.16	1.06	4.51	1.2	2.43	2.6	0.33
1.003	10.35	1.07	4.23	1.25	2.10	2.8	0.28
1.004	9.77	1.08	3.98	1.3	1.85	3.0	0.24
1.005	9.33	1.09	3.77	1.35	1.64	3.5	0.17
1.006	8.96	1.10	3.58	1.4	1.46	4.0	0.13
1.007	8.66	1.11	3.41	1.45	1.32	4.5	0.10
1.008	8.39	1.12	3.26	1.5	1.19	5.0	0.08
1.009	8.16	1.13	3.12	1.6	1.00	6.	0.06
1.01	7.95	1.14	3.00	1.7	0.86	7.	0.04
1.015	7.15	1.15	2.88	1.8	0.75	8.	0.03
1.02	6.59	1.16	2.78	1.9	0.66	9.	0.03
1.03	5.81	1.17	2.68	2.0	0.58	10.	0.02

It may be desirable, and it is equally easy, to analyze records directly in terms of the best fit to a Chapman distribution. The true height of maximum density, for example, may be found immediately as it is equal to the virtual height at the frequency for which $f/f_0=0.811$. A number of other ratios of f to f_0 for which there are simple relations between h_M, h', and H are listed in Table I. If the virtual heights are read from the record at the frequencies corresponding to the tabulated values of f/f_0, it is easy to determine both h_M and H. (Note that corrections to be described below must be applied in case of the existence of a retarding layer below the reflecting layer.) In general it will be found that the values of H determined from the larger values of f/f_0 may, for daytime records, be greater than the values found from the middle frequency range. This is an indication of a deviation from the Chapman distribution of densities, a deviation which is presumably caused either by imperfect superposition of a number of density curves, each of which may have the Chapman form, or by upward diffusion of free electrons.

By an approximate computation exactly similar to that described above, it is possible to determine the retardation of a ray that passes completely through an ionized layer having a

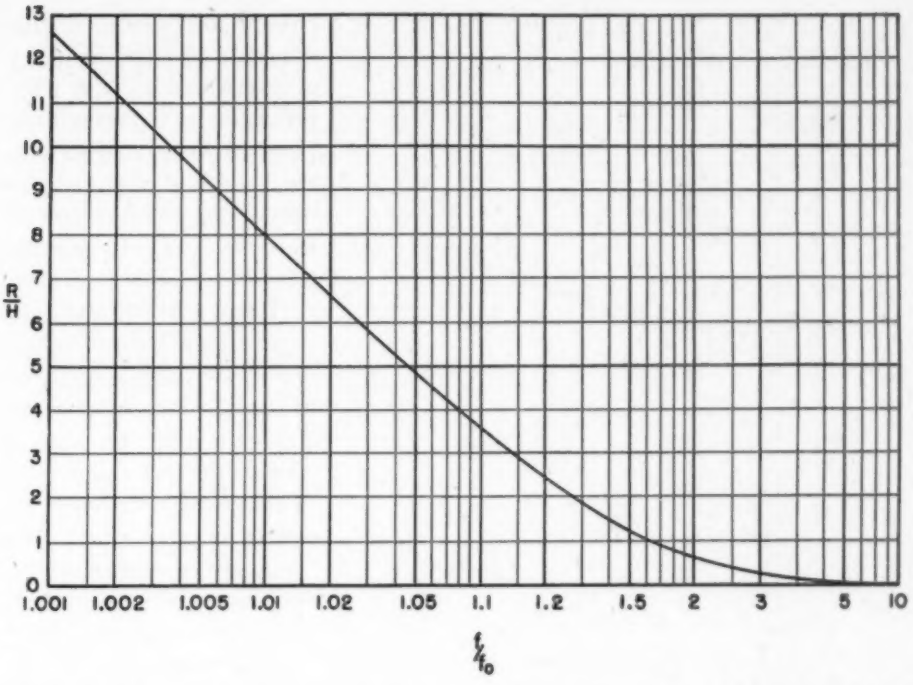


Fig. 5. The "retardation" of a ray which passes normally through a layer having a Chapman distribution.

Chapman distribution. These data are exhibited in Table II and in Fig. 5, which is drawn to a scale of f/f_0 which is linear in log (f/f_0-1). The quantity R/H is the "retardation," or the contribution to the total virtual height which is due to the decrease in group velocity in passage through the retarding layer, divided by the scale height for that layer. The data given assume that the retarded ray passes completely through the layer, but the figures are reduced by only a few percent if the ray passes no further than a point which is three or four times the scale height beyond the maximum.

Combining the data from Figs. 3 and 5, we find that, for a ray whose frequency is greater than the critical frequency for the lower layer and smaller than the critical frequency for the upper layer,

$$h'=h_{MF}+H_F\left[\frac{h'-h_M}{H}\right]_F+H_E\left[\frac{R}{H}\right]_E, \qquad (5)$$

where the subscripts F and E refer to the con-

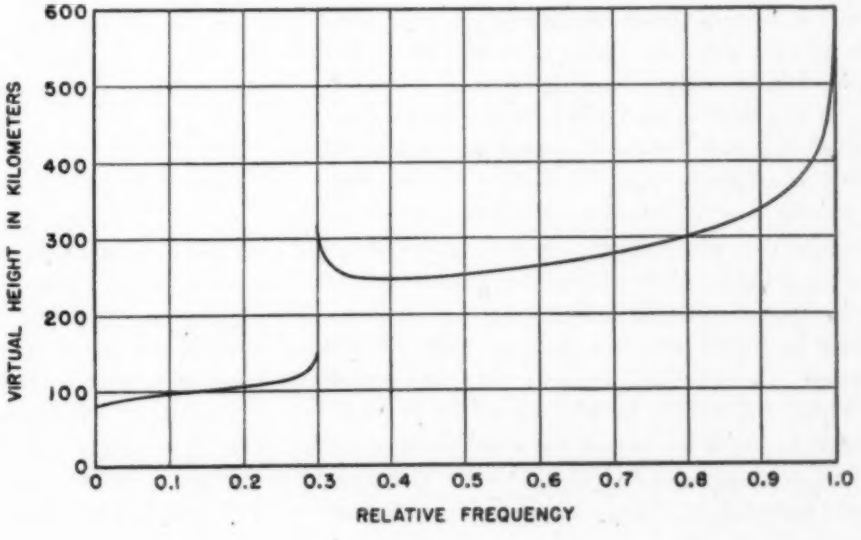


FIG. 6. Virtual height-frequency diagram for an ionosphere consisting of two Chapman layers.

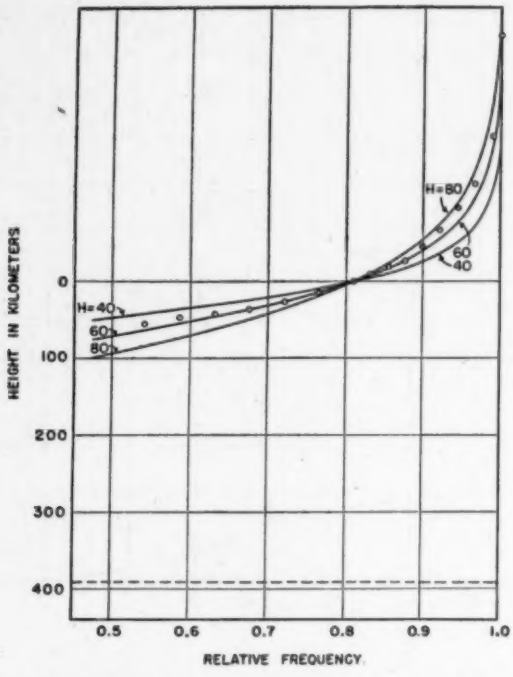


FIG. 7. An example of one method for analysing experimental virtual height-frequency data for true height and scale height.

stants for the upper and lower layers, respectively.

Figure 6 exhibits the variation of the virtual height with frequency for an hypothetical ionosphere having the following characteristics: $h_{MF}=300$ km; $H_F=50$ km; $h_{ME}=110$ km; $H_E=10$ km. Here the ratio of the critical frequencies for the E and F regions is 0.3. It will be noted that, unless the retarding layer overlaps the reflecting layer significantly, the virtual height for the upper layer is independent of the true height of the retarding layer.

It is probable that the best simple method of analyzing data for the F_2 region is to replot them to a standard scale of f/f_0 and to find by inspection the best fit between these data and one of a family of virtual height curves, calculated for various values of H, which have been drawn on a transparent template. The height of maximum may be found from the distance from the point where the curves intersect on the template (where $f/f_0=0.811$ in the case of no E retardation) to the base line of the experimental data. This procedure, of course, yields the true height and scale height of the Chapman layer that most closely corresponds to the actual distribution. A diagram of a rudimentary template of this sort is shown in Fig. 7, in which the circles represent points on an experimental curve supposedly seen through the transparent template, and in which the base line of the experimental data is shown dotted. In this case (00^h24^m on October 2, 1940) the F layer may be represented by the quantities: $h_M=392\pm5$ km; $H=65\pm5$ km; and $N_{max}=(2.46\pm0.01)10^5$ ions/cm^3.

At times when E-layer ionization is in evidence, the analysis may be made by preparing a number of templates of this sort from the data in Tables I and II, each template being for a specific ratio of E-layer critical frequency to that of the F region. When this has been done, it is simply necessary to select the proper template for each experimental curve before reading off the values of h_M and H. This procedure implies that H is known for the E layer.

In the presence of F_1 layer ionization the analysis is more difficult because the F_1 and F_2 layers overlap to such an extent that the virtual height for the F_2 layer cannot be obtained correctly by estimating F_1 retardation in the way in which it can be done for the E layer. It is still possible, of course, to trace out the total virtual height by the laminar method, but this process has not been carried to the point where a summary of the results can be expressed easily. A rough method for subtracting the effect of the F_1 region so that the constants for the F_2 region can be obtained is evident, however, from the construction of Figs. 8 and 9. In Fig. 8 the diagram at the top represents a variation of density of ionization with height obtained by adding three Chapman distributions, as follows:

E layer	F_1 layer	F_2 layer
$h_M=110$ km	$h_M=185$ km	$h_M=300$ km
$H=10$ km	$H=10$ km	$H=50$ km
$N_{max}=0.2\ N_0$	$N_{max}=0.25\ N_0$	$N_{max}=N_0$

At the bottom of the figure is the corresponding virtual height curve which was calculated by the laminar method. The dotted curve is the F_2 layer part of the figure as it would appear if the F_1 ionization were subtracted from the total.

Exactly the same construction is followed in Fig. 9, with the important exception that here the height of the F_1 maximum has been taken as 198 km. Under these conditions no subsidiary

maximum appears on the height-density curve and consequently there is no critical frequency for the F_1 layer.

In each case it is apparent that the dotted virtual-height curve, representing the F_2 layer in the absence of the F_1 layer, passes nearly midway between the minimum virtual height for the F_2 layer and the minimum virtual height for the F_1 layer. This clue enables us to obtain approximate values of the Chapman distribution constants for the F_2 layer. If the experimental virtual-height curve resembles the solid curve of Fig. 8 or Fig. 9, it is easy to sketch upon it a single smooth curve that will not differ from the curve which would be obtained from the F_2 layer alone by more than a few kilometers of virtual height. This curve may be compared with the templates described above in order to determine h_M and H for the F_2 layer, while these constants for the F_1 layer can be found by the treatment which has been described for two layers.

It is not within the scope of this article to discuss the presence of an E_2 layer or of other ionization between the E and F regions. Further study may yield a sufficiently accurate method of analysis for the retardation introduced by such ionization so that we can form some conclusions about its distribution.

It should be noticed that the assumption of no magnetic field leads, in the case of the ordinary ray which alone is considered here, to values of the virtual height that are always too large except for the case of transverse propagation. These errors, become smaller at the higher frequencies.

It may also be worth while to note that if the assumption of Chapman form for the various layers is at all justified, the construction of Fig. 8 indicates that the scale height for the F_1 layer must be small compared with that for F_2. It is evident that no F_1 maximum can occur unless the negative gradient of the F_1 ionization above the maximum is steeper than the positive gradient of the bottom of the F_2 layer. Since the curves are quite symmetrical near the maximum and have the same form, their slopes will have the same magnitude only if the scale heights for the two layers are approximately in the same ratio as their maximum densities of ionization. This argument, together with the known variations of the F_1 critical frequency, indicates that the F_1

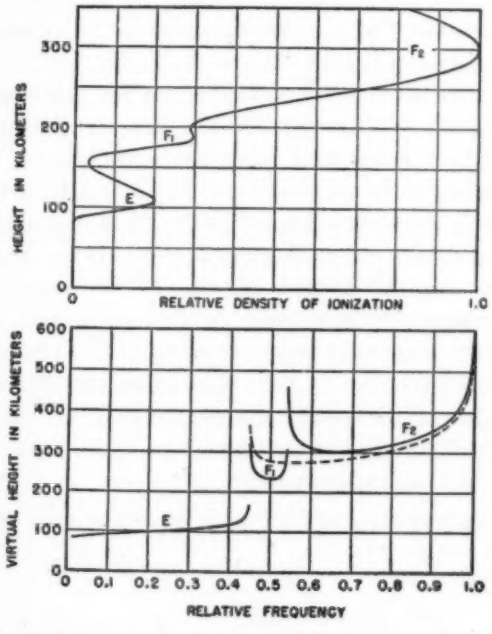


FIG. 8. An assumed distribution of electron densities in the ionosphere, and the corresponding virtual height-frequency diagram. The dotted curve would be obtained if the F_1 ionization were subtracted from the total.

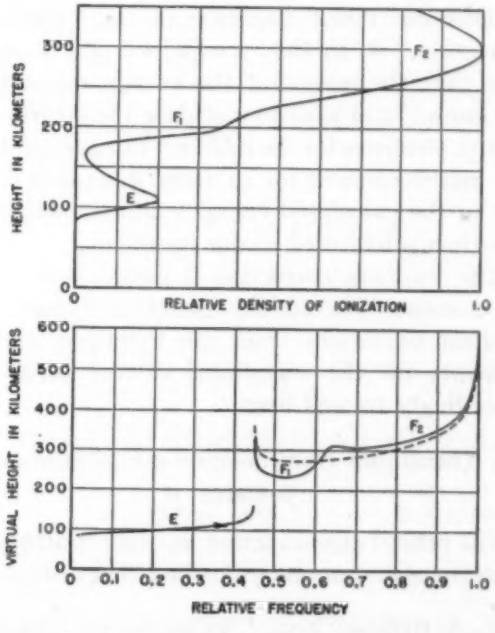


FIG. 9. Another assumed distribution of ionization with the corresponding virtual height-frequency curve. In this case the F_1 ionization is at such a height that no density maximum exists for that layer.

layer is very similar to the E in scale height, in maximum density of ionization, in the extent and symmetry of its diurnal variation, and perhaps in its behavior over longer periods. In case these ideas are correct, we must look for an explanation of the seasonal appearance and disappearance of the F_1 critical frequency in the distribution of ionization with height in the F_2 layer.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

The Calculation of the Coordinate Matrix Elements for Helium

WERNER ROMBERG
University, Oslo, Norway
(Received December 9, 1946)

Formulae for the calculation of transition probabilities in an atom with two electrons are derived for transitions from discrete s to both continuous and discrete p-states.

I. INTRODUCTION

IT is of interest, especially on account of its application to astrophysics, to investigate the influence which the mutual interaction of the two electrons in helium has on the matrix elements of the coordinates. The various perturbations, the coulomb interaction, the exchange, and the "polarization" all contribute to the energy. As a consequence, the chief quantum number, n, is no longer integral. These perturbations are taken together as the "Rydberg correction," δ, so that $n=[n]+\delta$. We assume here that the results of the energy calculation are known[1] and wish to calculate the coordinate matrix elements for variable n.[2] In order to find the matrix element for an atom, δ is taken from one of the calculated energies and the value of n is then substituted in our formulae.

The Rydberg correction is particularly large for s-states. We should therefore expect the greatest deviations from the hydrogen matrix elements for the transitions sp and these are accordingly treated here.

1. Transitions from discrete s to continuous p-states

The orbital eigenfunction Ψ, after splitting off from the spin function, must be either symmetric (para) or antisymmetric (ortho) in the electrons

$$\Psi=2^{-\frac{1}{2}}[\Psi_i(r_1)\Psi_a(r_2)\pm\Psi_i(r_2)\Psi_a(r_1)].$$

The inner electron is in the ground state

$$\psi_i=e^{-Zr/a},\quad a=\hbar^2/m_0e^2.$$

Z is the nuclear charge in the field of which the inner electron moves. For the outer electron moving in the reduced field z, we will assume that the eigenfunction Ψ_a may be expressed as the product of a radial function and a spherical harmonic function, $\Psi_a=\psi(r)\cdot Y_{lm}(\Theta,\Phi)$. If this is substituted in the expressions for the matrix elements of the coordinates, exactly the same selection rules and the same coefficients, depending only on l and m, as in the hydrogen matrix elements result.[3] Only the integrals depending on the radius are altered. If we denote with capital letters N, L, M the quantum numbers corresponding to the state with the higher azimuthal quantum number, $L=l+1$, and substitute for r the dimensionless quantity $\rho=2zr/a$, we obtain the expression $R_{nl}{}^{NL}\cdot a/2z$, which Gordon denoted by C.

For the (sp)-transitions

$$R_{nl}{}^{NL}=(II_{nN}\pm I_NI_n)\cdot[I_{NN}I(I_{nn}I\pm I_n{}^2)]^{-\frac{1}{2}}\quad(1,1)$$

where

$$I_{nN}=\int_0^\infty\psi_{n0}\psi_{N1}\rho^2d\rho;\quad I_{NN}=\int_0^\infty\psi_{N1}{}^2\rho^2d\rho;$$

[1] E. A. Hylleraas, Zeits. f. Physik 66, 453 (1930), 65, 209 (1930). H. A. Bethe, *Handbuch der Physik* [1], Vol. 24, p. 347.

[2] I have to thank E. A. Hylleraas for this suggestion. See also reference 5.

[3] W. Gordon, Ann. d. Physik 2, 1031 (1929).

$$I_{nn}=\int_0^\infty \psi_{n0}{}^2\rho^2 d\rho;\quad I=\int_0^\infty e^{-Z\rho/z}\rho^2 d\rho=2(z/Z)^3;$$

$$I_N=\int_0^\infty e^{-Z\rho/2z}\psi_{N1}\rho^3 d\rho;\quad I_n=\int_0^\infty e^{-Z\rho/2z}\psi_{n0}\rho^2 d\rho.$$

2. The approximate functions and their normalization

The difficulty lies in discovering suitable approximate functions for the discrete spectrum. In his earlier[4] calculation of the helium matrix elements, E. A. Hylleraas used the functions which the variation method yielded from the energy calculation. Lately,[5] Hylleraas has used complex integrals in order to extend the eigenfunctions for integral n, the Laguerre polynomials, to arbitrary n. This method shows itself to be especially suitable for the calculation of the transitions to the continuum.

The differential equation for ψ, if we neglect the polarization[6] and assume that the inner electron effectively reduces the nuclear charge to ze, is

$$\rho^2\psi''+2\rho\psi'-(l(l+1)-\rho+\rho^2/4n^2)\psi=0, \quad (2,1)$$

which is satisfied by

$$\psi(\rho)=\rho^l n^{-2l-1}(2\pi i)^{-1}\int e^{-\rho(1-2\zeta)/2n}\times(1-\zeta)^{l+n}\zeta^{l-n}d\zeta, \quad (2,2)$$

if the path of integration is taken so that at both ends

$$(1-\zeta)^{l+1+n}\zeta^{l+1-n}e^{-\rho(1-2\zeta)/2n} \quad (2,3)$$

has the same value.

For the discrete s-states, following Hylleraas, we take the path from $\zeta=-\infty$ round $\zeta=0$ back to $\zeta=-\infty$, whereby (2, 3) vanishes so long as $\rho>0$. If we let $\zeta=s/(1+s)$ and integrate by parts

$$\psi_n(\rho)=n(2\pi i\rho)^{-1}\int_{-\infty}^{(0,+)} e^{-\rho(1-2\zeta)/2n}\times(1-\zeta)^{n-1}\zeta^{-n-1}d\zeta$$
$$=n(2\pi i\rho)^{-1}\int_{-1}^{(0,+)} e^{-\rho(1-s)/(1+s)2n}s^{-n-1}ds. \quad (2,4)$$

ψ_n does not fulfill the boundary conditions at $\rho=0$, but can be used as an approximate function for the outer electron for the calculation of the matrix elements, as the deviation from the exact solution in the region $\rho=0$ contributes very little to the matrix integral I_{nN}.

The normalizing integral is easily found. It is

$$I_{nn}=2n^3\{1-[\Psi'(n)-(1/n)+(1/2n^2)]\times(\sin n\pi)^2/\pi^2\}. \quad (2,5)$$

The expansion of $\Psi'(n)=d^2/dn^2\log\Gamma(n)$ in reciprocal powers of n is[7] $\Psi'(n)=(1/n)-(1/2n^2)+(1/6n^3)-(1/30n^5)+\cdots$ so that the coefficient of $(\sin n\pi)^2$ is small compared with unity.

The interaction integrals for large, real quantum numbers are proportional to n^{-3} and vanish at the boundary of the continuum. The eigenfunctions of the continuum are hydrogen eigenfunctions but in the reduced field, z. We can take them from Sommerfeld,[8] remembering that his ρ is our ρ/n

$$\psi=\frac{\Pi(n+l)}{\Pi(n-l-1)}i^l n^{-l-1}R$$
$$=\frac{\Pi(n+l)}{\Pi(n-l-1)}\rho^{-l-1}(2\pi i)^{-1}\oint (x+\tfrac{1}{2})^{n-l-1}\times(x-\tfrac{1}{2})^{-n-l-1}e^{-\rho x/n}dx. \quad (2,6)$$

Sommerfeld gives for the normalizing integral

$$\int_0^\infty R^2r^2dr=\frac{\pi}{2}k_s{}^{-2}e^{-\pi|n|}|\Gamma(n+l+1)|^{-2}(an^2/z);$$
$$k_s=z/|n|a.$$

---

[4] E. A. Hylleraas, Zeits. f. Physik **106**, 395 (1937).
[5] E. A. Hylleraas, Archiv f. Math. og Natv. **48**, No. 4 (1945).
[6] We take account of the polarization approximately as it is included in $n=[n]+\delta$.
[7] Jahnke-Emde, *Funktionentafeln* (B. G. Teubner, Leipzig, 1938), p. 20.
[8] A. Sommerfeld, *Atombau und Spektrallinien II* (Friedrich Vieweg und Sohn, Braunschweig, 1938), p. 121, Eq. (28). The eigenfunction ψ is also given directly from our Eq. (2.2) by integrating by parts $(2l+1)$ times over the closed path about $\zeta=0$ and $\zeta=1$.

The last factor appears when we normalize for the quantum number interval $dn=1$.

The principal quantum number of the continuum is purely imaginary and we put $N=i\nu$, $\nu=(2m_0E)^{-\frac{1}{2}}hz/a$, where E denotes the energy of the free electron.

Hence, for the p-state

$$I_{NN}=2\nu(\nu^2+1)(1-e^{-2\pi\nu}) \qquad (2,7)$$

and with $x=\eta+\frac{1}{2}$

$$\psi_N=(N^2-1)N\rho^{-2}(2\pi i)^{-1}\oint_{(-1,0)}(1+\eta)^{N-2}\times\eta^{-N-2}e^{-\rho(1+2\eta)/2N}d\eta. \qquad (2,8)$$

3. The matrix integral I_{nN}

The integral I_{nN} is first integrated over ρ

$$I_{nN}=nN^2(N^2-1)(2\pi i)^{-2}\int_{-\infty}^{(0+)}(1-\zeta)^{n-1}\zeta^{-n-1}\times\oint_{(-1,0)}(1+\eta)^{N-2}\eta^{-N-2}(\eta-\eta_0)^{-1}d\eta d\zeta. \qquad (3,1)$$

The integral with respect to η is equal to the negative residue at the pole

$$\eta_0=-\tfrac{1}{2}+N(2\zeta-1)/2n.$$

The transformation $2\zeta=1-1/y$ together with the abbreviation

$$f=n\log\frac{1+y}{1-y}-N\log\frac{1+ky}{1-ky}$$

$$=n\log\frac{1+y}{1-y}-2\nu\arctan(ny/\nu);$$

$$k=n/N=n/i\nu \qquad (3,2)$$

gives

$$I_{nN}=2^5n^5(1-1/N^2)e^{i\pi\nu}(2\pi i)^{-1}\times\int_0^{(1+)}e^{f}y^4(1-y^2)^{-1}(1-k^2y^2)^{-2}dy. \qquad (3,3)$$

Integrating by parts

$$I_{nN}=-8n^2(1-k^2)^{-2}\{2(1-k^2)(\sin n\pi)/\pi+G_2+2nG_1+G_0\} \qquad (3,4)$$

where

$$G_0=-(\sin n\pi)/\pi,$$
$$G_j=e^{i\pi\nu}(2\pi i)^{-1}\int_0^{(1+)}y^{-j}\frac{de^f}{dy}dy;\quad j=0,1,2. \qquad (3,5)$$

This form of the integrand suggests the introduction of f as a new variable.[9]

In order to calculate G_j we deform the path of integration so that it passes through the regular point $y=\infty$; in fact, we put $y=re^{-i\pi/3}$ along the path from 0 to ∞ and $y=re^{+i\pi/3}$ along the path back to 0. Then

$$G_j=\frac{1}{\pi}\lim_{r\to\infty}Re\left\{ie^{-i\pi\nu}\int_0^{y=r\exp(i\pi/3)}y^{-j}\frac{de^f}{dy}dy\right\}. \qquad (3,6)$$

We expand f in powers of y and split up the coefficient of the lowest power, $f=-\frac{2}{3}n(1-k^2)\varphi$. For the integrals G_j we require y^{-1} and y^{-2} as power series in φ.

We let

$$y^{-1}e^{i\pi/3}=\frac{1}{r}=\sum_{p=0}^{\infty}a_p(e^{2i\pi/3}(1+k^2)/5)^p\varphi^{(2p-1)/3}$$

and

$$y^{-2}e^{2i\pi/3}=\frac{1}{r^2}=\sum_{p=0}^{\infty}b_p(e^{2i\pi/3}(1+k^2)/5)^p\varphi^{(2p-2)/3} \qquad (3,7)$$

where a_p, b_p are coefficients to be determined later. Then

$$G_1=\frac{1}{\pi}\sum_p^{\infty}a_p\Pi\left(\frac{2p-1}{3}\right)[(1+k^2)/5]^p\times[2n(1-k^2)/3]^{(1-2p)/3}\sin(-n-\tfrac{1}{3}+2p/3)\pi,$$

$$G_2=\frac{1}{\pi}\sum_{p=0}^{\infty}b_p\Pi\left(\frac{2p-2}{3}\right)[(1+k^2)/5]^p\times[2n(1-k^2)/3]^{(2-2p)/3}\sin(-n-\tfrac{2}{3}+2p/3)\pi. \qquad (3,8)$$

4. The calculation of the coefficients a_p, b_p

The most convenient relation for the calculation of a_p, b_p is

$$3\frac{d}{d\varphi}y^{-1}=(1-1/y^2)(k^2-1/y^2) \qquad (4,1)$$

[9] Method of steepest descent.

TABLE I. Table of coefficients from the recursion formula (4, 4).

	$1\cdot$	$+a_1\cdot$	$+(a_2)^2\cdot$	$+(a_2)^3\cdot$	$+(a_2)^4\cdot$	$+(a_2)^5\cdot$
$a_2=$	0	1				
$a_3=$	0.11111111	−0.22222222				
$a_4=$	−0.11111111	−0.14141414	−0.54545455			
$a_5=$	0	−0.35431235	−0.24941725			
$a_6=$	−0.01975309	0.13029440	0.00676854	0.16969697		
$a_7=$	0.03209877	0.08669704	0.60214615	0.44129988		
$a_8=$	0	0.15052357	0.21678623	0.48512744	0.09073510	
$a_9=$	0.00463779	−0.06659891	−0.02767115	−0.35543989	−0.17129693	
$a_{10}=$	−0.01039911	−0.04400100	−0.44506038	−0.70035860	−0.80378893	−0.14795090
$a_{11}=$	0	−0.06435739	−0.14151900	−0.63408920	−0.54310902	−0.27641329
	$1\cdot$	$+a_2\cdot$	$+(a_2)^2\cdot$	$+(a_2)^3\cdot$	$+(a_2)^4\cdot$	$+(a_2)^5\cdot$
$b_2=$	1	2				
$b_3=$	0.22222222	1.55555556				
$b_4=$	0	−0.72727273	−0.09090909			
$b_5=$	−0.22222222	−0.76923077	−2.03418803			
$b_6=$	−0.02716049	−0.71964085	−0.71874299	−0.75151515		
$b_7=$	0	0.45194022	0.45084327	0.96558344		
$b_8=$	0.07654321	0.38762433	2.14170974	2.13151380	0.81838480	
$b_9=$	0.00488601	0.34851639	0.65085535	1.95543443	0.91814678	
$b_{10}=$	0	−0.23956751	−0.44040667	−1.85259927	−1.16922010	−0.29955557
$b_{11}=$	−0.02793128	−0.18833842	−1.60090500	−3.14552213	−4.50545115	−1.79771769

which, together with

$$b_p=\sum_{q=0}^{p} a_q a_{p-q} \tag{4, 2}$$

give

$$a_0=1;\quad a_1=1;\quad a_2=(1-4k^2)(4-k^2)(1+k^2)^{-2}/7;$$
$$b_0=1;\quad b_1=2;\quad b_2=1+2a_2 \tag{4, 3}$$

and the following coefficients from the recursion formulae

$$(2p+3)a_p=5b_{p-1}-\sum_{q=1}^{p-1}(b_q b_{p-q}+2a_q a_{p-q});$$
$$b_p=2a_p+\sum_{q=1}^{p-1} a_q a_{p-q}. \tag{4, 4}$$

The next coefficients are given as polynomials in a_2 in Table I. The series (3, 8) for G_1 and G_2 are asymptotic series in which the terms decrease rapidly after the first. Only for very small ν, i.e., for high energies of the free electron, is a greater number of terms necessary.

5. The exchange integrals I_N and I_n

The integral I_N can easily be calculated, as the path of integration for η in (2, 8) is closed. We integrate for ρ, put $2\eta=y-1$ and $m=z/Z$ and integrate by parts.

$$I_N=-2^6N^3(N^2-1)(2\pi i)^{-1}\oint_{(-1,\,+1)}(y+1)^{N-3}(y-1)^{-N-3}(2y+N)(y-y_0)^{-1}dy. \tag{5, 1}$$

The path of integration can be contracted round the pole $y_0=-N/m=-i\nu/m$. Then

$$I_N=4^3N^4(N^2-1)(N-m)^{N-3}(N+m)^{-N-3}(m-2)m^5 \tag{5, 2}$$

or

$$I_N=4^3\nu^4(\nu^2+1)(\nu^2+m^2)^{-3}(m-2)m^5\exp[-2\nu\arctan(m/\nu)]. \tag{5, 3}$$

The integral I_n is integrated first over ρ and then by parts

$$I_n=4n^2m^2(n^2-m^2)^{-1}\Big\{[2n(1-m)(n+m)^{-1}-1](\sin n\pi)/\pi$$
$$+4mn^2(1-m)(n+m)^{-2}(2\pi i)^{-1}\int_{-1}^{(0+)}s^{-n}(1+st)^{-1}ds\Big\};\quad t=(n-m)/(n+m). \tag{5,4}$$

Expanding the denominator and integrating round a unit circle

$$I_n=4n^2m^2(n+m)^{-3}\left\{n+m-2mn+\frac{4m^2n}{1-n}+4mn^2(1-m)(n+m)^{-1}\sum_{p=1}^{\infty}t^{p-1}/(p+1-n)\right\}(\sin n\pi)/\pi. \quad (5,5)$$

This series converges rapidly only when t is small compared with 1. As $m=z/Z<1$, generally $n\gg m$. In order to find an expansion in powers of n^{-1} we transform the pole to 0 by putting $st=y-1$ and deform the path of integration so that it runs from $y=1-t$ in the negative imaginary half-plane to $y=-\infty$ and back in the positive imaginary half to $1-t$. Then

$$\int_{-1}^{(0+)}s^{-n}(1+st)^{-1}ds=-t^{n-1}\left(\cos n\pi\oint(1-y)^{-n}dy/y+2i\sin n\pi\int_{-\infty}^{1-t}(1-y)^{-n}dy/y\right).$$

The first integral is given by contracting both paths about the pole and has the value $2\pi i$. Substituting $y=x/(n+m)$ in the second integral

$$\int_{-\infty}^{1-t}(1-y)^{-n}dy/y=\int_{-\infty}^{2m}e^x/x\exp[-x-n\log(1-x/(n+m))]dx.$$

Expanding the second exponential function in powers of $1/n$ and integrating term by term, we have finally

$$I_n=-4m^2n^2(n^2-m^2)^{-1}(\sin n\pi)/\pi-16m^3n^4(1-m)(n^2-m^2)^{-2}t^n \times\{\cos n\pi+[\langle Ei\rangle_m(2m)+e^{2m}C](\sin n\pi)/\pi\},$$

$$C=-\tfrac{1}{2}m+(2m-1)/12n^2+(20m^4-4m^3+6m^2-6m+3)/90n^4+\cdots. \quad (5,6)$$

In the case of helium, $Z=2$, $m=z/Z=\frac{1}{2}$, so that

$$I_n=-4n^2(4n^2-1)^{-1}(\sin n\pi)/\pi-16n^4(4n^2-1)^{-2}t^n \times\{\cos n\pi+[\langle Ei\rangle_m(1)-e+(e/40n^4)+\cdots](\sin n\pi)/\pi\}, \quad (5,7)$$

to an accuracy of over 0.01 percent for all $n>1.5$.

II. TRANSITIONS FROM DISCRETE s- TO DISCRETE p-STATES

We wish to investigate the junction between the continuous and the discrete spectrum. We shall then be able to neglect the Rydberg correction for the discrete p-state in comparison with its much greater value for the s-state, so that N is a whole number. With this assumption E. A. Hylleraas[3] has calculated the matrix integrals. His series are, however, differently constructed so that an immediate comparison is not easily made. We shall therefore indicate shortly how our formulae for the discrete spectrum are to be altered.

As approximate function for the discrete p-state we use (2, 2) in which, as N is integral, the path of integration can be taken round $\zeta=0$ and $\zeta=1$. Integrating by parts three times gives (2, 8). The normalizing integral is

$$I_{NN}=2N(N^2-1). \quad (6,1)$$

At the boundary of the continuum $N\to\infty$, $\nu\to\infty$, Eqs. (6, 1) and (2, 7) become identical.

Equations (5, 1) and (5, 2) for I_N are always valid because $N>1>m$. Our expression for the matrix integral I_{nN} can be taken over unchanged so long as $N\gg n$.

The matrix elements of the continuum therefore join continuously with those for the discrete spectrum.

When $N \ll n$ the path of integration is taken along $y = re^{\pm 2\pi i/3}$. This gives the same formulae (3, 8) for G_j; instead of $(1-k^2)^{(j-2p)/3} \cdot \sin[-n-(j-2p)/3]\pi$ we now have

$$(k^2-1)^{(j-2p)/3} \cdot \sin[-n-(2j+2p)/3]\pi.$$

The asymptotic series (3, 8) for G_1 and G_2 become useless as N approaches n. We shall therefore, for the sake of completeness, also give expressions which can be used when $N \approx n$.

We substitute $y=(1+s)/(1-s)$ in Eq. (3, 5) and expand the integral in powers of $\epsilon=(N-n)/(N+n)$. Then

$$G_2+G_0=2n\epsilon(A+B)(\sin n\pi)/\pi; \quad G_1=n\epsilon(A-B)(\sin n\pi)/\pi$$

and

$$I_{nN}=16n^2N(1+k)^{-3}\{(1+k)^2+n(A+B)+n^2(A-B)\}(\sin n\pi)/(n-N)\pi,$$

where

$$A=-(N+n)^{-1}S_0{}^1+\sum_{p=1}^{N} S_{p-1}{}^0\epsilon^{p-1}/(N-n-p)+\sum_{p=1}^{\infty} S_0{}^{p+1}(-\epsilon)^{p+1}/(N-n+p),$$

$$B=+(N+n)^{-1}S_1{}^0+\sum_{p=1}^{N-2} S_{p+1}{}^0\epsilon^{p+1}/(N-n-p)+\sum_{p=1}^{\infty} S_0{}^{p-1}(-\epsilon)^{p-1}/(N-n+p).$$

For

$$S_a{}^b=\sum_{q=0}^{N-1-a}\binom{N-1}{a+q}\binom{N+b+q}{N}(-\epsilon^2)^q,$$

we have the following recursion formulae

$$(N+b)\cdot[S_0{}^{b-1}+\epsilon^2 S_0{}^{b+1}]=S_0{}^b\cdot[(1+\epsilon^2)b+2N\epsilon^2],$$

$$(N-a)\cdot[S_{a-1}{}^0+\epsilon^2 S_{a+1}{}^0]=S_a{}^0\cdot[(1+\epsilon^2)a-2N\epsilon^2],$$

and

$$S_0{}^1-S_1{}^0=2S_0{}^0, \quad S_{N+p}{}^0=S_N{}^0=0; \quad S_{N-1}{}^0=1.$$

In order to calculate transitions from discrete p- to continuous s-states the approximate function for the p-state must be changed, as also for transitions with higher l. The matrix integral I_{nN} can be calculated as the corresponding integrals G_j can be reduced to our G_1 and G_2 by integration by parts. But the approximate function can then no longer be normalized.

Investigation and also numerical calculations along these lines are in progress.

I wish to thank Professor E. A. Hylleraas at Oslo, Norway, for his most valuable encouragement in this work. I should also like to thank him and Professor I. Waller at Uppsala, Sweden, for enabling me to carry it out during a critical time.

The Mechanism of the Negative Point Corona at Atmospheric Pressure in Relation to the First Townsend Coefficient

LEONARD B. LOEB
Department of Physics, University of California, Berkeley, California
(Received February 11, 1947)

On the basis of recent work by P. L. Morton and G. W. Johnson on ionization by electrons in divergent electrical fields, it is shown that at atmospheric pressure it is permissible to compute the ionization in electron avalanches by means of the Townsend expression $e^{\int \alpha dx}$ down to points of the order of 1×10^{-3} cm radius. The phenomena resembling a glow discharge, observed at atmospheric pressure with the negative points normally used, must be ascribed to ionization of the Morton type which can only be produced in the *positive space charge fields* built up at the point *later* in the course of pulse formation. The ionization at both positive and negative points must initially be of the Townsend type. This conclusion clarifies difficulties, and is consistent with the pulse build-up time of 0.4 microsecond just observed by W. N. English with a Synchroscope. An explanation is also given for the equality of starting potentials of pulses from positive and negative points in air, while in *pure* N_2 and H_2, the negative pulses start at much lower potentials.

THE work of Trichel,[1] Weissler,[2] and of Loeb, Kip, and Hudson[3] on the negative point corona in air and other gases at atmospheric pressure has established a close analog between its character and structure and that of the well-known glow discharge. This behavior, in contrast to the phenomena observed with a positive point corona under similar conditions, leads to inconsistencies in the earlier interpretations given.[1] The difficulty was introduced through the use of the function $e^{\int \alpha dx}$ in calculating electron multiplication in the divergent fields about gaps, with α representing Townsend's first coefficient for ionization by electron impact, and x, the distance traversed in the field. The function yields a maximum of ionization adjacent to the cathode surface.[1] On the publication of Morton's[4] important study concerning the ionization by electrons in highly divergent fields, the difficulty became more acute. Morton showed that the Townsend function could not be applied to fields whose divergence was such that the field changed by more than 2 percent over an electron free path. Not only did Morton show that the ionization in such cases could reach values of an order of magnitude more than given by $e^{\int \alpha dx}$, but he further showed that under these conditions the ionization *near* the negative point electrode was very small, becoming appreciable only at some distance from the point. Such a distribution of ionization is just the inverse of that given by $e^{\int \alpha dx}$. Thus the Morton process does not yield the conventional Townsend electron avalanches. Morton's findings are in excellent accord with the situation required to account for the observed mechanism near the cathode of a glow-like discharge. However, they introduce some difficulty when it is noted that the use of $e^{\int \alpha dx}$ nicely accounts for the streamer and other phenomena at a *positive* point of moderate diameter,[5] while the Morton conditions seem to apply to the *negative* corona from the same point, at fields of the same magnitude.

A comprehensive extension of Morton's initial studies over a pressure range from 0.01 mm to 760 mm pressure in pure H_2 and mercury free air has just been completed by G. W. Johnson[6] in the writer's laboratory, and will be published shortly. On the basis of the new data the writer was able to make some calculations which enable the difficulty properly to be clarified. It appears that at *atmospheric pressure* for all gases studied, the geometrical field conditions for concentric cylinders are such that the electrons remain *pretty well in equilibrium* in the electrical fields down to wire diameters of the order of 1×10^{-3}

[1] G. W. Trichel, Phys. Rev. **54**, 1078 (1938).
[2] G. L. Weissler, Phys. Rev. **63**, 96 (1943).
[3] L. B. Loeb, A. F. Kip and G. G. Hudson, Phys. Rev. **60**, 714 (1941).
[4] P. L. Morton, Phys. Rev. **70**, 358 (1946).
[5] K. E. Fitzsimmons, Phys. Rev. **61**, 175 (1942).
[6] G. W. Johnson, Phys. Rev. **71**, 278(A) (1947).

cm. With other geometrical configurations, such as hemispherical capped cylinders or confocal parabaloids,[7] the value of the critical point radius might be slightly larger, but not materially so. Thus, the use of $e^{\int \alpha dx}$ in calculation for most corona points commonly used at atmospheric pressure is justified, unless *space charge conditions* alter the fields materially.

This also justifies previous practice in computing the creation of the avalanche leading to streamer formation in *positive* point corona, which determines onset.[5] In such calculations, theory and experiment have always been in satisfactory agreement.[5,6]

In the case of the *negative* corona the calculations for electron multiplication leading to evaluation of the onset threshold must extend beyond the first ionizing avalanches to which $e^{\int \alpha dx}$ can apply; for, unlike the streamer, or the burst pulse, the negative corona pulse builds up more gradually.* In this instance, at the threshold, the first electrons leaving the negative point region move outward into lower fields producing $e^{\int \alpha dx}$ electrons and positive ions as the static field requires. Photons produced by this avalanche, together with some positive ion impact on the negative point, then produce more avalanches, and a continuous or chain succession of such events builds up. During this period, however, the relatively slowly moving positive ions are approaching the point, building up a notable positive ion space charge at the point which is only partially neutralized by the rapidly moving secondary electrons and their progeny leaving the point. In consequence, there is built up a *very strong potential gradient at the cathode surface* which is *many times the geometrically fixed gradient starting the phenomenon*. There is nothing in the corona environment to prevent the development of cathode gradients from becoming quite comparable to those observed in glow discharges. When such gradients are achieved, in conformity with the Morton findings, the electrons then gain some twenty to a hundred electron volts of energy in ten to twenty electron free paths in the dark space. With ionization and excitation probabilities of the order of a few percent these electrons lose little energy. They thus do not ionize very much until they reach the weak field region beyond positive space charge. This mechanism yields the characteristic Crookes dark space and negative glow so often described.[1,2,3] Using $e^{\int \alpha dx}$, without the building up of a space charge and the Morton condition, as in the past, this observed condition was hard to understand.[1] Such a discharge will continue indefinitely with some fluctuations in magnitude until choked off by negative ion formation and space charge,[3] or until alteration of the cathode surface by positive ion bombardment reduces secondary emission below the operating threshold.[2] The mechanism indicated at once removes the conflict between the positive corona, operating with $e^{\int \alpha dx}$, and the negative corona operating as a glow discharge, in that, initially, $e^{\int \alpha dx}$ applies to both points but alters to Morton's condition as the discharge develops. The clarification, however, does more in that it makes possible an understanding of sputtering[2,3] observed in negative coronas. With the positive space charge gradients the positive ions can strike the cathode with ten or more volts of energy instead of the estimated 1 volt following on the Townsend calculation.[3]

TABLE I. Data on the potential in volts for the beginning of bursts for positive and negative points ($Vg\pm$) and for the *onset* phenomena of Weissler ($V_{onset}\pm$).

Gas	$Vg+$	$Vg-$	Self-sustaining corona $V_{onset}+$	Self-sustaining corona $V_{onset}-$
H_2	3500	2800	—	—
N_2	4800	3700	—	—
Air (Weissler)	—	—	5500	8000
Air (English)	5100±50	5000±50	5550±50	8000–10,000 Indefinite oscillographically, individual pulses still resolved

It had been observed by Weissler that in *very pure* H_2 and N_2 the corona starting potentials were considerably lower for the negative than for the positive point. On the other hand, it has been generally asserted, without data being published, that the starting potentials for positive and negative points in air was about the same and higher than in pure H_2 and N_2. This is illustrated by the data of Table I. These data

[7] L. H. Fisher and G. L. Weissler, Phys. Rev. 66, 102 (1944).

* Recent studies of the Trichel negative corona pulses in air by means of a Synchroscope with high sweep rate by W. N. English in this laboratory indicate that the unresolved pulses of Trichel do indeed build up gradually, rising to a peak and declining over a time interval of 0.4 microsecond, while streamers build up more rapidly.

represent the beginning of bursts for the positive points, $Vg+$, and Trichel pulses, $Vg-$, for the negative points as shown by oscillograph for air, together with the corresponding data for the *onset* phenomena observed by Weissler in N_2 and H_2. Weissler failed to specify his point diameter for the data given. He observed the onset visually and by current change for the *self-sustaining* coronas in H_2, N_2, and air. In this discussion the potentials for the *beginning* of pulses, Vg only is of interest. W. N. English kindly has measured the potentials for room air by using a 0.5 mm point. This, it appears, was also the point that Weissler used. The values in volts are:

An explanation of this marked difference in behavior was indicated by Weissler,[2] but needs further clarification in the light of what has been stated above. There appear to be two factors present to effect such a change. First, in pure gases the secondary mechanisms responsible for onset of the *negative* corona are determined by secondary electron liberation by photoelectric effect from the cathode, by positive ion bombardment of the cathode, or by photoelectric ionization in the gas near the cathode. In the *positive* corona the fundamental secondary mechanism active is *only* photoelectric ionization in the gas leading to streamer, or burst pulse, production. Secondly, in air, electrons of low energy *do not remain free* but form negative ions of heat, of formation in the order of 0.2 volt.[3] In pure H_2 and N_2 electrons are free. Thus, from the viewpoint of initiating electrons, the pure gases merely require adequate fields to initiate secondary mechanisms near the point. The secondary liberation of electrons from a negative point in *pure gases* requires a relatively low field, since the photoelectric effect from the *cathode metal* is adequate in the relatively transparent gases to furnish such electrons at the start. With the discharge well developed, secondary electron liberation by positive ion impact on the cathode is also very effective. In the pure gases the photoelectric ionization *in the gas* is very meager, and it is not surprising to see that higher fields may be required for a discharge operating by this action only, in the case of the positive point in pure H_2 and N_2. In air, on the other hand, the heterogeneous character of the gas facilitates photoelectric ionization in the gas, so that the positive threshold could be lowered relative to that in pure N_2 and H_2, provided that Townsend's first coefficient is not materially altered by O_2. On the other hand this circumstance *lowers* the efficiency of photoelectric ionization *at the cathode* by gas absorption of the photons. It can also raise the work function of the cathode because of the action of oxygen on the surface. It is, however, hardly likely that at any time the *photoelectric action in the gas* can be the dominating element in negative point corona, though it, together with photoelectric action at the cathode, may help start the discharge in air. After space charges build up, positive ion bombardment doubtless predominates. Thus, in regard to secondary mechanism, the negative corona should usually have a lower threshold. However, neither of these discharges in air can start until the fields near the electrodes are sufficiently high so that a field strength to pressure ratio, X/p, exceeding 90, the value required to liberate electrons from the negative O_2 ions,[3] extends far enough out from the point to insure adequate electron ionization in the avalanches. This situation appears to be achieved in air at the starting potential for preonset streamers and burst pulses expected in positive corona.[5] Thus, it is this circumstance that probably fixes the value for the onset of *both* the *negative* and *positive* corona pulses at about the same potential, under conditions where, without this difficulty, the negative starting point could well be lower.

[3] L. B. Loeb, Phys. Rev. **48**, 684 (1935). See also N. E. Bradbury and G. Bloch, Phys. Rev. **48**, 689 (1935).

Direct Determination of Stacking Disorder in Layer Structures

W. H. ZACHARIASEN
Department of Physics, University of Chicago, Chicago, Illinois
(Received February 14, 1947)

A stack of N identical, parallel, and equidistant layers is used as a simple model of a layer structure. Irregularities in the relative displacements $\boldsymbol{\delta}$ parallel to the plane of the layers correspond to stacking disorder. The stacking disorder is described by means of a set of functions $W_M(\boldsymbol{\delta})$ which measure the probability of finding a relative displacement $\boldsymbol{\delta}$ for layers M spacings apart. It is shown that x-ray diffraction data permit a direct determination of the Fourier coefficients of the functions W_M. Thus these functions can be synthesized and the nature of the stacking disorder directly deduced.

INTRODUCTION

IRREGULARITIES in the relative displacement of the layers parallel to their own planes occur in many crystals of layer structure type. This stacking disorder gives rise to characteristic features in the x-ray and electron diffraction patterns of such crystals. The purpose of this paper is to show that the specific nature of the stacking disorder can be directly deduced from the observed diffraction effects.

In order to demonstrate clearly the general principles upon which the method is based it is convenient to use a simple model of a layer structure. The extension of the method to layer structures of more general type can be made without serious difficulty.

A SIMPLE MODEL OF A LAYER STRUCTURE

Our model is a sequence of N identical, parallel, and equidistant layers. The periods within the layers will be denoted by $\mathbf{a}_1$ and $\mathbf{a}_2$, while $\mathbf{a}_3$ is the period in the third principal direction (not necessarily normal to the plane of the layers). The individual layers in the stack will be numbered consecutively by means of an index L_3, the layer $L_3=0$ being chosen arbitrarily. The symbol $\boldsymbol{\delta}_{L_3}$ will be used to represent the displacement, parallel to the plane of the layers, of layer L_3 relative to layer $L_3=0$.

The position vector of any atom in the crystal has the form

$$\mathbf{r}_j+L_1\mathbf{a}_1+L_2\mathbf{a}_2+L_3\mathbf{a}_3+\boldsymbol{\delta}_{L_3}.$$

$\mathbf{r}_j$ is the position vector of the jth atom within a unit parallelogram of a layer and L_1, L_2, L_3 are a set of three integers.

The model just described corresponds to an ordered layer structure if $\boldsymbol{\delta}_{L_3}$ varies periodically as the index L_3 is allowed to run through the sequence of integral value. We shall assume that the $\boldsymbol{\delta}$-values are not periodic, in which case the model corresponds to a layer structure with stacking disorder.

We shall assume that the distribution of $\boldsymbol{\delta}$-values throughout the stack is independent of the crystal surface. In order to study the consequences of this assumption we imagine the stack of layers to be periodically repeated in space. Our assumption implies that any sequence of N consecutive layers bodily removed from the periodic stack corresponds to a crystal specimen representative of the given disorder. Clearly there are N such different but statistically equivalent specimens. In discussing the disorder we shall therefore have to think in terms of the assembly of N specimens rather than in terms of a single specimen.

The stacking disorder for the assembly can be described by means of a series of probability functions $W_M(\boldsymbol{\delta})$, where $M=0, 1, 2, \cdots, N-1$. A function W_M represents the probability of finding a relative displacement $\boldsymbol{\delta}$ between two layers M spacings apart.

The displacement $\boldsymbol{\delta}$ can be expressed in the form $\boldsymbol{\delta}=y_1\mathbf{a}_1+y_2\mathbf{a}_2$ where $0\leq y_i<1$. Clearly $W_0=0$ for all values y_1, y_2 except $y_1=y_2=0$. Since the functions W_M are even and periodic functions we may set

$$W_M(y_1, y_2)=\sum_{H_1}\sum_{H_2} W^M{}_{H_1H_2}\cos 2\pi(H_1y_1+H_2y_2), \quad (1)$$

where the Fourier coefficients conversely are

given by

$$W^M{}_{H_1H_2}=\int_0^1\int_0^1 W_M \cos 2\pi(H_1y_1+H_2y_2)dy_1dy_2.$$

THE X-RAY DIFFRACTION PATTERN

This section is devoted to a theoretical study of the x-ray diffraction patterns of structures with stacking disorder. Since each individual layer is assumed to be a fully ordered two-dimensional lattice we shall obviously find that the scattering is restricted to directions with "Miller indices" (H_1H_2z) when H_1 and H_2 (but not necessarily z) are integers. We shall show that the intensity of scattering in a direction (H_1H_2z) can be expressed in terms of the Fourier coefficients $W^M{}_{H_1H_2}$. Secondly we shall show that the individual Fourier coefficients can be deduced from the experimental data and that hence the probability functions W_M can be synthesized.

The mean intensity of the scattered x-rays from the assembly of N crystals discussed in the preceding section becomes

$$I(\mathbf{s})=NA(\mathbf{s})\sum_M \cos\mathbf{s}\cdot M\mathbf{a}_3\langle\cos\mathbf{s}\cdot\boldsymbol{\delta}_M\rangle_{Av}. \quad (2)$$

In this expression $\mathbf{s}=2\pi(\mathbf{k}-\mathbf{k}_0)$ where $\mathbf{k}_0$ and $\mathbf{k}$ are the wave vectors of incidence and of scattering. $A(\mathbf{s})$ represents the intensity of scattering from a single layer.

It is convenient to set

$$\mathbf{s}=2\pi(H_1\mathbf{b}_1+H_2\mathbf{b}_2)+2\pi(p_1\mathbf{b}_1+p_2\mathbf{b}_2+z\mathbf{b}_3)$$

where H_1 and H_2 are integers and $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ the vector set reciprocal to $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$. In this way the functions $I_{H_1H_2}(p_1p_2z)$ represent the intensity distribution in the reciprocal lattice.

The function $A_{H_1H_2}(p_1, p_2\, z)$ for given z has a maximum at $p_1=p_2=0$. The half width of this maximum depends upon the lateral dimensions of the layers. Unless the crystal dimensions parallel to the layers are of colloidal or subcolloidal order of magnitude one will not be able to observe the variation of $A_{H_1H_2}$ as function of p_1 and p_2. For this reason we shall deal not with the intensity $I_{H_1H_2}(p_1p_2z)$ but with the experimentally significant quantity $P_{H_1H_2}(z)$ defined by

$$P_{H_1H_2}(z)=\int\int I_{H_1H_2}(p_1, p_2, z)dp_1dp_2.$$

The expression for $P_{H_1H_2}$ becomes

$$P_{H_1H_2}(z)=I_0\delta VB_{H_1H_2}(z)\sum W^M{}_{H_1H_2}\cos 2\pi Mz. \quad (3)$$

The quantities $W^M{}_{H_1H_2}$ are the Fourier coefficients which were introduced in the preceding section. I_0 is the intensity of the incident beam and δV is the volume of a single crystal specimen, which for the sake of convenience is assumed to be so small that absorption and extinction can be neglected. The function $B_{H_1H_2}$ can be written in the form

$$B_{H_1H_2}=V^{-1}T|\psi_{H_1H_2}|^2$$

where $V=\mathbf{a}_1\cdot\mathbf{a}_2\times\mathbf{a}_3$, $\psi_{H_1H_2}$ the structure factor and $T=(1+\cos^2 2\theta)e^4/2m^2c^4R^2$ the Thomson scattering factor.

Experimentally one does not measure directly the power distribution $P_{H_1H_2}(z)$ in the reciprocal lattice, but rather the power $Q_{H_1H_2}(\alpha)$ in terms of a parameter α which is suitable for the particular experimental procedure which is used. The reduction of the observations $Q_{H_1H_2}(\alpha)$ to the reciprocal lattice distribution $P_{H_1H_2}(z)$ is, however, a trivial problem which involves the evaluation of the appropriate Lorentz factor.

Since $W^M{}_{00}=1$ for any M the function $P_{00}(z)$ will have maxima for integral values of $z=H_3$, and the half width of these maxima is determined by the total height of the stack of layers. In other words the stacking disorder has no effect on the reflections $(00H_3)$.

If the nature of the stacking disorder, i.e., each function W_M, is known, the calculation of the functions $P_{H_1H_2}(z)$ by means of Eq. (3) is a simple matter, and the nature of the stacking disorder can be found by a trial and error procedure. Equation (3) suggests, however, a direct method for the determination of the specific character of the disorder.

The distributions $P_{H_1H_2}(z)$ can be regarded as functions which are known from experiments, however, on a relative rather than absolute scale. The functions $B_{H_1H_2}(z)$ are known if the structure of each layer has been determined.

Hence, the quantities

$$D_{H_1H_2}(z)\equiv P_{H_1H_2}/I_0\delta VB_{H_1H_2}$$

are known functions except for a numerical scale factor. By the Fourier reciprocity theorem we

have

$$W^{M}{}_{H_1H_2}=\int_{-\frac{1}{2}}^{\frac{1}{2}} D_{H_1H_2}(z)\cos 2\pi Mz\,dz.$$

Accordingly the probability functions W_M which characterize the stacking disorder can be synthesized. The unknown numerical scale factor can be determined from the requirement that W_M is everywhere positive and $\int\int W_M dy_1 dy_2 = 1$.

In most instances it is found that the functions W_M approach a limit W_∞ as M increases. It is therefore convenient to set

$$W_M = W_\infty + \Gamma_M$$

where accordingly Γ_M represents fluctuations at small values of M so that the Fourier coefficients $\Gamma^{M}{}_{H_1H_2}$ go towards zero with increasing M.

One sees readily that

$$\sum_M W^{M}{}_{H_1H_2}\cos 2\pi Mz = W^{\infty}{}_{H_1H_2}\sum_M \cos 2\pi Mz + \sum_M \Gamma^{M}{}_{H_1H_2}\cos 2\pi Mz.$$

Thus the functions W_∞ is generally to be associated with the sharp maxima, the functions Γ_M with the diffuse maxima of the functions $P_{H_1H_2}$.

If the function W_∞ is a constant, implying zero correlation between layers far apart, $W_{H_1H_2}=0$ unless $H_1=H_2=0$. Apart from the reflections $(00H_3)$ all maxima of $P_{H_1H_2}$ will then be diffuse.

The writer has used the method outlined above to determine the stacking disorder in several crystals of the layer structure type.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

Surface States and Rectification at a Metal Semi-Conductor Contact

JOHN BARDEEN
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received February 13, 1947)

Localized states (Tamm levels), having energies distributed in the "forbidden" range between the filled band and the conduction band, may exist at the surface of a semi-conductor. A condition of no net charge on the surface atoms may correspond to a partial filling of these states. If the density of surface levels is sufficiently high, there will be an appreciable double layer at the free surface of a semi-conductor formed from a net charge from electrons in surface states and a space charge of opposite sign, similar to that at a rectifying junction, extending into the semi-conductor. This double layer tends to make the work function independent of the height of the Fermi level in the interior (which in turn depends on impurity content). If contact is made with a metal, the difference in work function between metal and semi-conductor is compensated by surface states charge, rather than by a space charge as is ordinarily assumed, so that the space charge layer is independent of the metal. Rectification characteristics are then independent of the metal. These ideas are used to explain results of Meyerhof and others on the relation between contact potential differences and rectification.

INTRODUCTION

THE generally accepted view[1] of the nature of the rectifying contact between a metal and a semi-conductor is illustrated in Fig. 1 which applies specifically to an excess semi-conductor. Figure 1a shows an energy level diagram of the metal and semi-conductor in equilibrium, but with the contact separated. The Fermi level is the same in both the metal and the semi-conductor. As shown, the work function of the metal, χ_1, is greater than the work function of the semi-conductor, χ_2^0, so that there is a contact difference in potential, $\chi_1-\chi_2^0$. It is assumed that when the metal and semi-conductor are nearly joined, the potential distribution is as shown in Fig. 1b. A double layer is formed such as to give a potential drop, φ_0, from the metal to the interior of the semi-conductor equal to the contact potential difference.

This double layer is assumed to consist of a space charge region in the semi-conductor, extending to a depth of the order of 10^{-6} to 10^{-4} cm,

[1] See, for example, N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1940), Chap. V. The theory of rectification is due in large part to W. Schottky. The most important of his papers is Zeits. f. Physik 118, 539 (1942).

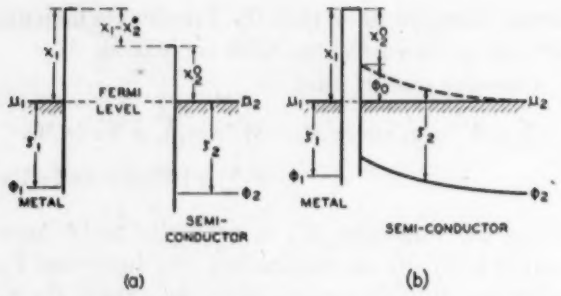


FIG. 1. Energy level diagram for metal and semi-conductor in electrical and thermal equilibrium. μ_1 and μ_2 represent the Fermi levels in metal and semi-conductor, respectively, and ϕ_1 and ϕ_2 the inner potentials. ζ_1 and ζ_2 are the chemical potentials and χ_1 and χ_2^0 the work functions. Figure 1a shows a wide gap with a contact potential difference $x_1-x_2^0$. In 1b, the gap is very small. There is a space charge region in the semi-conductor which gives an electrostatic potential energy rise at the surface $\varphi_0=\chi_1-\chi_2^0$. This is the usual picture in which no surface states on the semi-conductor are assumed.

and an induced charge on the metal surface. The space charge gives a rise in electrostatic potential energy at the surface of the semi-conductor. Electrons are depleted from the space charge region, giving a layer of high resistivity. If a potential is applied to the junction, most of the drop occurs across this barrier layer. If the potential of the semi-conductor is negative with respect to the metal, the electron energy levels in the semi-conductor are raised, and electrons may flow more easily over the potential hill into the metal. This is the direction of easy flow. On the other hand, if the semi-conductor is positive, the levels are lowered, increasing the height of the hill, and making it more difficult for electrons to travel from the semi-conductor to the metal. This is the direction of high resistance.

According to this view, the equilibrium height of the potential hill, φ_0, and therefore also the degree of rectification, depend on the work function of the metal. For an excess semi-conductor, the larger the work function of the metal, the larger is the potential rise, and the larger is the reverse resistance of the barrier. If the work function of the metal is less than that of the semi-conductor there is a potential drop instead of a potential rise, and no rectification will occur. For a defect semi-conductor, in which the current is carried by holes, just the reverse is true: low metal work function gives high rectification.

A number of investigations have been carried out which have verified these conclusions in some cases, and in other cases have not. H. Schweickert, as quoted by Schottky,[2] has found a correlation between the resistance of selenium rectifiers in the blocking direction and the work function of the metal. Selenium is a defect conductor, and high reverse resistance was found for low work function metals such as K, Na, Li, Ba, and low reverse resistance was found for such high work function metals as Ag, Au, Ni. The metal electrodes were put on by evaporation.

W. H. Brattain,[3] working in this laboratory, has found a good correlation between degree of rectification and work function for metal contacts evaporated on cuprous oxide (a defect conductor) and on both N- and P-type silicon.[4] Metals used, listed in order of decreasing degree of rectification, were Al, Ag, and Pt on cuprous oxide; Pt, Be, Ag, Mg, and Al on N-type silicon; and Mg, Cd, Ag, and Pt on P-type silicon. He found that when contact is made to the semi-conductor by a metal junction in air, the rectification is practically independent of the work function of the metal used. Results somewhat similar to those of Schweickert have been obtained by J. N. Shive[3] who studied the rectification characteristics of a number of contacts made by evaporation of various metals on selenium. Metal contacts, listed in order of decreasing degree of rectification, are Be, Zn, Pb, and Au.

A. V. Joffe[5] has studied contact potential differences and the resistances of the contacts formed from a large number of different semi-conductors and metals. Most gave very poor rectification characteristics. While there was some qualitative correlation of contact potential differences with contact resistance, quantitative agreement with theories of Schottky[1] and Davydov[6] was poor.

---

[2] W. Schottky, Physik Zeits. 41, 570 (1940).

[3] Unpublished work done at the Bell Telephone Laboratories in 1940 under the general direction of J. A. Becker. The author is indebted to Drs. Brattain and Shive for permission to quote their results.

[4] The designations N- and P-type refer to the direction of rectification. An N-type is an excess semi-conductor, and the direction of easy flow occurs when the semi-conductor is negative relative to the metal. A P-type semi-conductor is a hole conductor, and the direction of easy flow is opposite. In both cases, the direction of easy flow is that in which the carrier moves from the semi-conductor to the metal.

[5] A. V. Joffe, J. Phys. USSR 10, 49 (1946).

[6] B. Davydov, J. Phys. USSR 4, 355 (1941).

W. E. Meyerhof[7] has recently been making an extensive study of the relation between contact potential difference and rectification for metal point contacts applied to Si and Ge. He has determined the potential rise, φ_0, from the rectification characteristics at different temperatures and has also measured the ordinary contact potential difference, or Volta potential, by a modified Kelvin bridge method. Preliminary results show little correlation between the two sets of measurements. In fact, he has found that φ_0 is practically independent of the work function of the metal for metal point—Si rectifiers. The same metal may rectify with both P- and N-type silicon, in opposite directions.

These negative results indicate that a closer analysis of the nature of the contact between a metal and semi-conductor is warranted. The main purpose of the present paper is to investigate the effect of electronic states on the surface of the semi-conductor on φ_0. There has been considerable discussion in the literature concerning the possibility of surface states,[8] but little direct evidence as to their existence.

IMPURITY LEVELS AND SURFACE STATES

According to the modern theory of semi-conductors, there is an energy gap between the highest filled band of levels and the lowest state of the conduction band. In the language appropriate to excess semi-conductors, conductivity results from electrons thermally excited to the conduction band from impurity levels. The impurity levels, which are intermediate in energy between the filled band and conduction band, represent states in which the electron is localized around a foreign atom or other defect in the crystal lattice. In normal semi-conductors there may be the order of 10^{-7} to 10^{-3} impurity levels for each atom of the crystal.

In addition to the impurity levels in the interior of the crystal, there may be localized states on the surface with energies in the "forbidden" region between the filled and conduction bands. Shockley[8] and others have investigated the conditions under which surface levels may be expected on an ideal crystal. His analysis, based primarily on a one-dimensional model, indicates that "in a plot of the energy spectrum *versus* interatomic distance the surface levels appear only at lattice constants so small that the boundary curves of the allowed energy bands have crossed." The number of such surface states is equal to the number of surface atoms, and, in a neutral crystal, the surface states are half filled. The conditions for surface states are fulfilled, for example, in a diamond-type lattice.

Surface states may also result from surface imperfections, from foreign atoms on the surface, etc. On general grounds, there is good reason to suppose that the ratio of the number of surface levels to surface atoms may be much higher than the ratio of the number of impurity levels to atoms in the interior.

The energy levels corresponding to the surface states may be discrete, or they may have a continuous distribution with all energies in the gap between the filled and conduction bands. There is little evidence of either an experimental or a theoretical nature on this point. Well-defined impurity states of the same nature, sparsely distributed over the surface, would all have about the same energy and form a discrete level. If the impurity atoms are densely distributed over the surface, so that there is considerable interaction between them, a continuous distribution in energy is to be expected. A continuous distribution is also expected for surface states on clean surfaces of the type discussed by Shockley, Goodwin, and others.[8] In the analysis to follow we assume that the surface levels are continuously distributed in energy. Many of the same general conclusions would follow, however, if discrete levels had been assumed.

We will show that if there is a relatively high density of surface states, there may be a double layer on the free surface of a semi-conductor re-

[7] W. E. Meyerhof, "Contact potential difference in crystal rectifiers," Technical Report No. 5, Univ. of Pennsylvania, BuShips contract NObs-34144, Aug. 10, 1946; see Phys. Rev. 71, 727 (1947).

[8] Surface states with energies in the "forbidden" band were studied first by I. Tamm, Physik. Zeits. Sowjetunion 1, 733 (1932), and are often called Tamm states. Later theoretical work was done by R. H. Fowler, Proc. Roy. Soc. A141, 56 (1933), S. Rijanow, Zeits. f. Physik 89, 806 (1934), A. W. Maue, Zeits. f. Physik 94, 717 (1935), E. T. Goodwin, Proc. Camb. Phil. Soc. 35, 205 (1939), W. G. Pollard, Phys. Rev. 56, 324 (1939), and W. Shockley, Phys. Rev. 56, 317 (1939). The last of these is the most critical, and shows under what conditions surface states are occupied in a normal crystal.

sulting from a surface charge caused by electrons in surface states and a space charge of opposite sign extending into the crystal to a depth of 10^{-6} to 10^{-4} cm. The double layer tends to make the work function independent of the impurities in the interior. If contact is made to a metal, the contact potential difference is compensated largely by a true surface charge rather than by space charge, so that the height of the barrier is largely independent of the metal.

NATURE OF CONTACT

Before discussing the effect of surface states, we will give a brief review of the nature of the contact between metal and semi-conductor.

At a contact between any two electronic conductors there is an electric double layer which adjusts the potential of one relative to the other for equilibrium conditions. This potential difference may be taken as a measure of the strength of the double layer.

Some care must be taken in defining the absolute value of the potential. There is, of course, no way of measuring the potential inside a metal or semi-conductor. The potential just outside the surface can be measured, but the difference between the potential inside and outside the surface depends on the double layer at the surface. Only differences in potential between different parts of a conductor can be observed directly.

Nevertheless, it is desirable to consider the potential in the interior. The true time-average electrostatic potential is periodic, with the period of the lattice. The macroscopic potential, Φ, may be defined as the space average of the actual potential over a region large compared with atomic dimensions. This definition is somewhat arbitrary; another definition might yield a different value for the strength of the double layer at a contact or at a surface. However, only differences in double layers have direct significance, and these are independent of the way the potential in the interior is defined.

The general theory of equilibria[9] shows that when two electronic conductors are in contact and in thermal equilibrium, the electrochemical potentials, μ_1 and μ_2, must be the same in both:

$$\mu_1=\mu_2 \text{ in equilibrium.} \tag{1}$$

In the usual picture of an electronic conductor, μ is just the Fermi level which determines the probability, p, that a state of energy E is occupied.

$$p=1/(1+\exp[(E-\mu)/kT]). \tag{2}$$

The electrochemical potential in a region may be defined thermodynamically, in a way which is independent of any particular model, by the equation:

$$\mu=\frac{\partial}{\partial n}(U-TS),$$

where n is the number of electrons in the region, and U and S are the energy and entropy. The partial differentiation corresponds to a reversible change in which the volume and temperature are held constant. The internal energy U includes an electrostatic potential energy, $-e\Phi$, per electron. The choice of the zero or reference level from which this potential energy is measured is arbitrary, and there is the same arbitrary choice in the definition of μ.

It is convenient to introduce a quantity which depends on the chemical constitution of the material, and on the electron density, but not on the electrostatic potential. This quantity, ζ, which is called the chemical potential, to distinguish it from the electrochemical potential, is defined by

$$\zeta=\mu-\varphi, \tag{3}$$

where

$$\varphi=-e\Phi$$

is the electrostatic energy per electron.

Let φ_s be the potential energy of an electron just outside the surface of a conductor. This potential is to be evaluated at a point where the image potential is negligible, but at a distance from the surface small compared with macroscopic dimensions. The work function of the surface is then:

$$\chi_s=\varphi_s-\mu. \tag{4}$$

This is the energy required to take an electron from the conductor and place it at rest at a point, defined as above, just outside the surface.

[9] An excellent discussion is given by R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, England, 1939), Chap. XI. The author is indebted to Dr. C. Herring for the method of presentation used above.

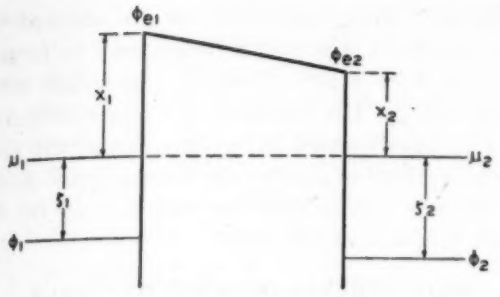


FIG. 2. Diagram to illustrate double layers at a contact between two conductors in equilibrium. There is a double layer of strength $\varphi_{e1}-\varphi_0$ at the surface of conductor 1, a double layer of strength $\varphi_{e2}-\varphi_2$, in the opposite direction at the surface of conductor 2, and a double layer of strength $\varphi_{e2}-\varphi_{e1}=-(\chi_1-\chi_2)$ formed by surface charges on the two conductors. The total strength, $\varphi_2-\varphi_1$, is independent of the surface double layers.

The work function depends on the double layer at the surface. By replacing μ by $\zeta+\varphi$, according to Eq. (3), the work function may be expressed in the form:

$$\chi_e=\varphi_e-\varphi-\zeta, \tag{5}$$

which shows explicitly that the work function is the sum of the energy required to take an electron through the surface double layer, $\varphi_e-\varphi$, and a body term, $-\zeta$, independent of the surface.

It should be emphasized that the separation of the work function into a surface term and a body term depends on the precise way the potential energy, φ, in the interior of the conductor is defined. We have suggested above that the electrostatic potential Φ be defined as space average of the actual potential. Another definition, might yield, for example, a smaller value for $\varphi_e-\varphi$ and a correspondingly larger value for $-\zeta$. Differences in double layers, such as might yield different work functions for different crystal faces of the same material, are not affected by the arbitrary definition of the inner potential. Since the double layer involves a potential difference, the value is independent of the reference level, or zero of potential, which is also arbitrary.

If two electronic conductors are in contact, and in thermal equilibrium, the electrochemical potentials must be the same in both. Thus,

$$\zeta_1+\varphi_1=\zeta_2+\varphi_2,$$

or

$$\varphi_2-\varphi_1=\zeta_1-\zeta_2. \tag{6}$$

The strength of the double layer at the contact, $\varphi_1-\varphi_2$, is equal to the difference between the chemical potentials. Since the chemical potentials depend only on the internal constitutions of the two conductors, the double layer is independent of the work functions of the two surfaces before they are brought into contact.

The situation when there is a small gap between the two conductors in equilibrium is illustrated in Fig. 2. There is a double layer of strength $\varphi_{e1}-\varphi_1$ at the surface of conductor 1, a double layer of strength $\varphi_{e2}-\varphi_2$, in the opposite direction, at the surface of conductor 2, and a double layer of strength $\varphi_{e2}-\varphi_{e1}=-(\chi_1-\chi_2)$, formed by surface charges on the two conductors. The total strength of the double layer is:

$$(\varphi_{e1}-\varphi_1)+(\varphi_{e2}-\varphi_{e1})-(\varphi_{e2}-\varphi_2)=\varphi_2-\varphi_1. \tag{7}$$

The surface double layers drop out, as they should according to Eq. (6).

If the surfaces are in very intimate contact, so that the electron space charges of the two surfaces overlap, it is not possible to divide the total double layer into different parts which depend on the work functions of the individual surfaces. All that can be said is that the strength of the total double layer is $\varphi_2-\varphi_1$. The work functions of the surfaces before they are brought into contact play no role.[10]

USUAL PICTURE OF RECTIFYING CONTACT

If one of the conductors is a semi-conductor, and the other is a metal, it has generally been assumed that as they are brought into contact, no surface charge forms on the semi-conductor.[11] Instead, there is a space charge near the surface which may extend to a depth of the order 10^{-6} to 10^{-4} cm. If the gap is small compared this distance, but is still large enough so that there is

[10] H. Y. Fan, Phys. Rev. 61, 365 (1942); 62, 388 (1942), has attempted to calculate the electron space charge distribution in the double layer at the contact between two metals, and at the contact between a metal and a semi-conductor. Although his calculations are based on a highly idealized model, they may serve to give a rough picture of the charge density in the contact region. He does not consider the effect of surface states on the semi-conductor.

[11] N. F. Mott, Proc. Camb. Phil. Soc. 34, 568 (1938) gives a discussion of the contact between a metal and an insulator or a semi-conductor. A brief analysis is given of the effect of a surface charge resulting from electrons in surface states on the contact between a metal and an insulator. Mott assumes a discrete energy level for the surface states.

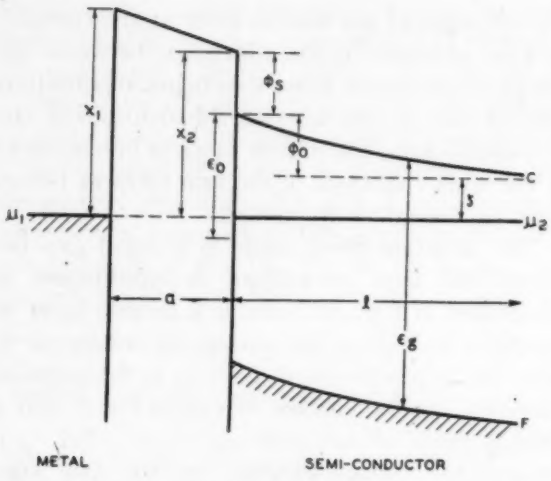


FIG. 3. Energy level diagram for metal semi-conductor contact illustrating notation used in text. The Fermi levels are μ_1 and μ_2 and the work functions χ_1 and χ_2. The lowest state of the conduction band is denoted by *C* and the highest level of the filled band by *F*. If the surface states are filled to an energy ϵ_0 below the conduction band there is no net charge on the surface atoms.

no appreciable overlap, most of the energy drop $\chi_1-\chi_2^0$ will occur in the semi-conductor rather than in the gap. This situation is illustrated in Fig. 1b. The space charge raises the energy at the surface of the semi-conductor by an amount $\chi_1-\chi_2^0$. This is the usual picture which indicates that the height of the potential barrier should be equal to the difference in work functions, and is that which is discussed in the introduction.

This picture is deficient in two respects:

(a) The contact may be so intimate that a division of the double layer into one at the surface of the metal, one at the surface of the semi-conductor, and one caused by space charge may not be possible. There will be a double layer at the immediate interface, and another due to space charge, but it is not possible to say how the total strength is divided between them. A cruder way of stating this is that the work functions of the two surfaces are modified by the contact. However, if the contact is intimate, the separate work functions have no meaning.

(b) There may be electronic states localized on the surface of the semi-conductor, so that the surface atoms can become charged. An appreciable field may then exist in the gap. Part of the total drop from metal to semi-conductor will occur across the gap and part across the space charge region. If the density of surface states is sufficiently high, part of the double layer at the *free* surface of a semi-conductor may be formed by a surface states charge compensated by a space charge. The height of the potential barrier of a rectifier formed from the material will then be determined in part by the normal space charge of the free semi-conductor and in part by the work function of the metal.

FREE SURFACE OF SEMI-CONDUCTOR

We will first consider the free surface of a semi-conductor, and then discuss the rectifying contact. The notation to be used is illustrated in the energy level diagram of Fig. 3. The lowest state of the conduction band and the highest state of the filled band of the semi-conductor are indicated, with an energy gap ϵ_g. It is assumed that the distribution of surface states is such that the surface states give no net charge if the states are filled to an energy ϵ_0 below the conduction band.[12] Since the Fermi level cuts the surface above the level determined by ϵ_0, the surface as shown will be negatively charged, this charge resulting from electrons in states between ϵ_0 and the Fermi level. The picture applies to an excess semi-conductor.[13] In the body of the

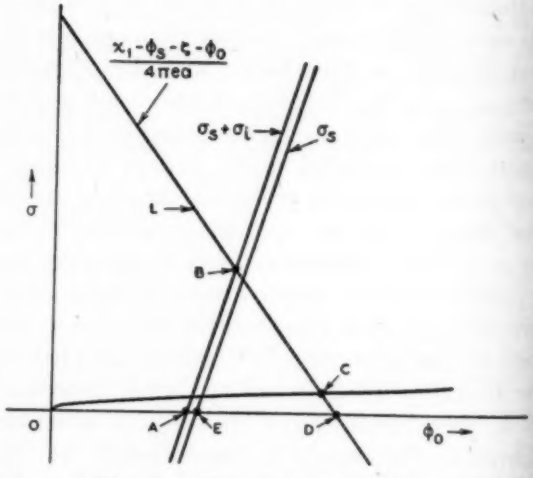


FIG. 4. Schematic diagram showing how the potential rise ϕ_0 is determined from the density of surface levels and the contact potential difference (see text).

[12] We refer to the net charge of the surface states as "surface states" charge to distinguish it from the total charge, including space charge, in the surface layer, which we call "surface charge."

[13] The case of an excess semi-conductor seems to be easier to visualize than that of a defect semi-conductor. All results derived for one case, of course, apply to the other with obvious changes in signs of the charges.

semi-conductor, the Fermi level is an energy ζ below the conduction band. The space charge region extends for an approximate distance l into the semi-conductor, giving a potential energy rise φ_0 at the surface. An energy φ_s is required to remove an electron from the lowest state of the conduction band near the surface to a point just outside the semi-conductor. The work function, χ_2, depends on φ_s as well as on φ_0:

$$\chi_2 = \varphi_s + \varphi_0 + \zeta. \tag{8}$$

The amount and extent of the space charge inside the free surface is determined by the density of surface levels. For zero external field, the positive space charge raises the potential at the surface by an amount just sufficient to give a compensating negative surface charge. The larger the positive space charge region, the larger is φ_0, and the smaller is the negative surface states charge. For some φ_0 the two will be equal in magnitude. This is the equilibrium value.

These relations are indicated in a schematic way in Fig. 4, which shows the variation of σ_i, the total space charge per unit area, of σ_s, the surface states charge, and of the surface charge, $\sigma_s+\sigma_i$, with φ_0. The other curves on the diagram will be referred to later in connection with the discussion of a metal contact. For a uniform space charge in the boundary layer, σ_i is proportional to the square root of φ_0 (see the appendix). The plot of σ_s has a steep slope corresponding to a fairly high density of surface states. It passes through zero (point E in Fig. 4) when

$$\varphi_0 = \epsilon_0 - \zeta, \tag{9}$$

because, according to the definition of ϵ_0 the surface states will then be filled up to the level corresponding to zero surface states charge. The point A represents a value of φ_0 for which $\sigma_s+\sigma_i=0$, which corresponds to a neutral surface. It can be seen that if the density of surface states is high, this value of φ_0 will be close to that given by Eq. (9).

For the limiting case of a very high density of surface states, the line σ_s becomes vertical, points A and E coincide, and Eq. (9) must be satisfied; hence, according to Eq. (8):

$$\chi_2 = \varphi_s + \epsilon_0. \tag{10}$$

The work function is determined entirely by the surface, and is independent of the position of the Fermi level in the interior.

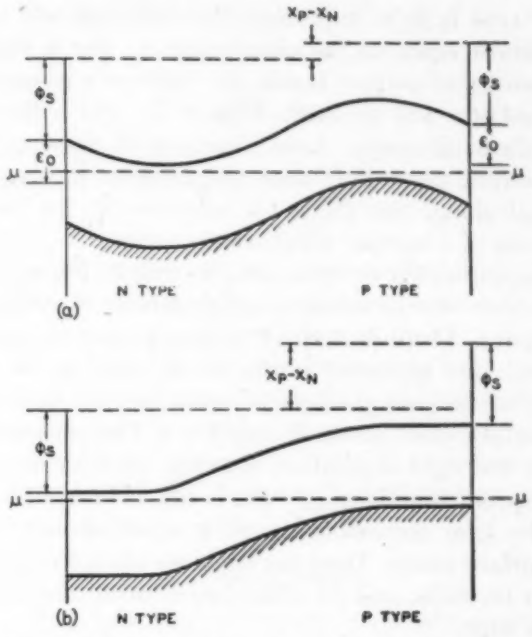


FIG. 5. Energy level diagram showing contact potential difference between two ends of a semi-conductor which changes from N-type to P-type along its length. (a) High density of surface states, small contact potential difference. (b) No surface states, large contact potential difference.

On the other hand, if the density of surface states is small, $\sigma_s+\sigma_i$ is approximately equal to σ_i and the condition for a neutral surface leads to a small value for φ_0. In the limiting case of vanishing surface states charge, $\varphi_0=0$, and the work function is

$$\chi_2 = \varphi_s + \zeta, \tag{11}$$

which of course *does* depend on the position of the Fermi level in the interior of the semi-conductor.

Some semi-conductors can be made either excess or defect, or N- or P-type,[4] depending on the nature and concentration of impurities. The distribution of impurities in a single sample may be such as to make one end N-type and the other end P-type. Silicon and germanium are examples of materials which may behave in this way. For an N-type conductor, ζ is small; for a P-type conductor ζ is almost equal to the energy gap ϵ_g, which is the order of one electron volt for these materials.

If the density of surface states is low, a large difference in work function between N- and

P-type is to be expected. The difference will be almost equal to the energy gap ϵ_g. For a high density of surface states, the difference in work function will be small. Figures 5a and b show schematic energy level diagrams of these two limiting cases. Both show the variation in potential along, and the work functions at the two ends of a sample which changes from N- to P-type from left to right along its length. Figure 5a is that corresponding to a high density of surface states. There is a space charge region at each end. The potential at the left is raised so as to bring the energy level corresponding to a neutral surface close to the Fermi level. The potential at the right is similarly lowered. The difference in work functions, $\chi_P - \chi_N$, is small. Figure 5b is the case corresponding to a small density of surface states. There are no space charge regions at the ends, and the difference in work functions is large.

A difference in work function can be detected experimentally as a difference in contact potential. Meyerhof[7] has in this way measured the differences in work functions of various samples of N- and P-type silicon and has investigated the effect of surface treatment on these differences. He finds that the difference between N- and P-type samples is about 0.25 ev, which is considerably smaller than the energy gap of about 1.1 ev. It is quite possible that there is a high density of surface states, either on the pure material or resulting from surface films or impurities, which is sufficient to account for this discrepancy.

We have so far discussed the density of surface states in a purely qualitative way. It is of interest to make an estimate of the density required to produce an appreciable space charge layer at the free surface of a semi-conductor. The density depends on the density of charged centers, or donors, in the boundary layer. If there are N such centers per unit volume, and if the thickness of the boundary layer (assumed uniform) is l, we have for the total space charge per unit area:[1]

$$\sigma_i = eNl. \tag{12}$$

The potential energy rise at the surface is

$$\varphi_0 = (2\pi/\kappa)e^2N l^2, \tag{13}$$

where κ is the dielectric constant.

Let the number of surface states in the energy interval $d\epsilon$ be $nd\epsilon/\epsilon_0$. We assume, for simplicity that n is a constant, which is of the order of the total number of surface levels per unit area with energies in the gap. The change in surface states charge density corresponding to a potential energy rise φ_0 at the surface is then

$$\Delta\sigma_s = en\varphi_0/\epsilon_0. \tag{14}$$

The condition for an appreciable space charge layer is that φ_0 be of the order of magnitude of ϵ_0 when $\Delta\sigma_s = \sigma_i$. This requires that n be the order of Nl. Taking, for example, $\kappa \sim 15$, and $N \sim 10^{17}$, Eq. (13) gives $l \sim 10^{-5}$, so that

$$n \sim Nl \sim 10^{12} \text{ states/cm}^2. \tag{15}$$

This density corresponds to about one surface state per thousand surface atoms. Other things being fixed, the limiting density of surface states is proportional to the square root of the density of charged centers in the boundary layer.

METAL SEMI-CONDUCTOR CONTACT

As a metal surface approaches that of a semi-conductor, the electrostatic field in the gap increases. There is a surface charge σ_M on the metal surface, and a charge of equal magnitude and opposite sign divided between a surface charge, σ_s, and a space charge, σ_i (the latter has been defined as the total space charge per unit area of surface). Thus

$$\sigma_M = -(\sigma_s + \sigma_i). \tag{16}$$

In the case of the metal, the surface charge causes only a very slight change in work function. However, the charge on the semi-conductor may cause appreciable changes in its work function, so that its value χ_2 will differ appreciably from the zero field value χ_2^0. These changes can be analyzed with the aid of Figs. 3 and 4.

According to Fig. 3, the field in the gap between metal and semi-conductor is $(\chi_1 - \chi_2)/ea$. This is produced by the surface charge which gives a field $4\pi(\sigma_s + \sigma_i)$. Equation (8) expresses χ_2 as a function of φ_0. Equating the two expressions for the field leads to:

$$\frac{\chi_1 - \zeta - \varphi_s - \varphi_0}{4\pi ea} = \sigma_s(\varphi_0) + \sigma_i(\varphi_0). \tag{17}$$

The left-hand side of the equation is plotted in Fig. 4 as the line "L," and is a straight line of negative slope. The solution to (17) is represented by the point B, which is the intersection of "L" with the line representing $\sigma_s+\sigma_i$. The line "L" crosses the axis at the point D, which corresponds to $\chi_2=\chi_1$, or

$$\varphi_0=\chi_1-\varphi_s-\zeta. \tag{18}$$

It can be seen that if the density of surface levels is high, so that σ_s and $\sigma_s+\sigma_i$ have steep slopes, the point B will lie close to the point A which gives the value of φ_0 for the free surface. In this case,

$$\varphi_0\sim\epsilon_0-\zeta, \tag{19}$$

and is practically independent of the work function of the metal. This is a possible explanation of the results of Meyerhof, who found that the value of φ_0 for metal points on silicon does not depend very much on the metal used.

In the limiting case of vanishing surface states charge density, $\sigma_s=0$, and the solution of (17) is given by the point C of Fig. 4. The value of φ_0 is then close to that given by the intersection of the line L with the horizontal axis (point D). In this case,

$$\varphi_0\sim\chi_1-\varphi_s-\zeta, \tag{20}$$

and the usual picture of the contact between metal and semi-conductor applies.

As the metal is brought closer to the semiconductor, the slope of the line "L" increases with decreasing "a" and one might expect a transition from a solution near A to one near D with a resulting value of φ_0 given by Eq. (20). However, even if there is only one surface state per hundred surface atoms, the slope of the line σ_s is so great that the solution still lies near point A even when a is reduced to atomic distances, say 3A. This may be verified by making a rough calculation based on Fig. 4. For the value of φ_0 to be largely determined by the semiconductor, the value of φ_0 for the solution B must lie closer to A then to D. The necessary and sufficient condition for this is that the slope of the line σ_s be much greater, in absolute value, than the slope of "L." We have

$$\begin{aligned}&\text{slope of } \sigma_s=m_s=en/\epsilon_0,\\ &\text{slope of } L=-m_L=-1/4\pi ea.\end{aligned} \tag{21}$$

The condition that m_s/m_L be large is

$$m_s/m_L=4\pi e^2an/\epsilon_0\gg 1, \tag{22}$$

or

$$n\gg\epsilon_0/4\pi e^2a. \tag{23}$$

Setting for example, $\epsilon_0\sim10^{-12}$ erg and $a\sim3\times10^{-8}$ cm, this condition requires that

$$n\gg10^{13}. \tag{24}$$

Thus if there is appreciably more than about one surface state per hundred surface atoms, the metal work function will have little effect on φ_0. This is about an order of magnitude larger than the surface density required for the existence of an appreciable boundary layer at the free surface.[14]

The theory is worked out in detail for the special case of a uniform Schottky exhaustion layer[1] in the appendix.

CONCLUSIONS

If the density of surface levels with energies in the "forbidden" band is sufficiently high ($>\sim10^{12}$/cm^2), there will be a double layer at the free surface of a semi-conductor formed from a surface states charge and a space charge of opposite sign. The space charge region is similar to that which exists at a rectifying contact. This double layer tends to make the work function independent of the height of the Fermi level in the interior, and so independent of the impurity content.

The total strength of the double layer at a rectifying junction between a metal and semiconductor is fixed by the difference in chemical potentials, and so depends on the body properties of the metal and semi-conductor, and is independent of the work functions of the surfaces before they are brought into contact. The double layer consists of the following parts:

[14] S. Benzer, Phys. Rev. 71, 141 (1947), has recently reported on the current-voltage characteristic observed when contact is made between two pieces of the same (homogeneous) germanium crystal. He states that "the characteristic observed for both polarities is more the order of the back resistance observed when either piece of the crystal is contacted with a metal; in both directions the negative resistance at high voltage appears, which is typical of the back characteristics of metal-Ge contacts using these alloys." The presence of a space charge layer at the surface of each piece which is little modified by contact would result in a characteristic similar to that of two rectifiers in series opposition. Such boundary layers are to be expected if the density of surface states is the order of that given by Eq. (24) and if the contact is not too intimate. Benzer's results are thus indirect evidence for surface states on germanium crystals.

(1) A double layer of atomic dimensions at the metal surface.

(2) A double layer of atomic dimensions at the semi-conductor surface.

(3) A double layer formed from surface charges on the metal and semi-conductor, both of atomic dimensions.

(4) A double layer formed from a surface charge of atomic dimensions and a space charge extending to a depth of 10^{-6} to 10^{-4} cm into the semi-conductor.

The strengths of the double layers may be estimated in different cases as follows:

(a) If the density of surface levels is sufficiently high ($>\sim 10^{13}/\text{cm}^2$) the double layer (4) will be independent of the metal, and will be the same as that for the free surface of the semi-conductor. The rectification properties will then be largely independent of the work function of the metal. The difference in contact potentials is compensated by (3).

(b) If the density of surface levels is small ($<\sim 10^{13}/\text{cm}^2$), the double layer (3) will be small, and (4) will be determined approximately by the difference in work functions.

(c) If the contact between the metal and semi-conductor is very intimate, it may not be possible to distinguish between the double layers (1), (2), and (3). The metal will tend to broaden the surface levels, but if the broadening is small compared with the energy gap, conclusion (a) will still be valid.

(d) If the broadening of the surface levels by the metal is large, no conclusions about the space charge can be drawn from measurements of contact potential differences.

It is believed that all of these cases may be realized.

The author is indebted to various members of the technical staff of the Bell Telephone Laboratories, particularly to W. Shockley, C. Herring, and W. H. Brattain, for the benefit of numerous discussions concerning the subject matter of this paper, and for helpful comments on the preparation of the manuscript.

APPENDIX

Theory for Schottky Exhaustion Layer

The detailed calculation of φ_0 requires a knowledge of the distribution of space charge in the semi-conductor and its dependence on φ_0. The dependence of surface charge on φ_0 is also required. We will carry through the calculation explicitly only for the case of a Schottky exhaustion layer[1] with a uniform density of charge. This case is particularly simple and brings out the essential features of the problem. For simplicity we also assume that the surface states are uniformly distributed in energy.

Let eN be the positive charge density in the barrier region, assumed constant. Let l be the thickness of the charged layer. Let $nd\epsilon/\epsilon_0$ be the number of surface levels per unit area with energies in the range $d\epsilon$. We assume that n is a constant, independent of energy. For other notation see Fig. 3.

We have the following relations:

(a) The total space charge per unit area is:

$$\sigma_s = Nel. \tag{12}$$

(b) The potential energy at the surface of the semi-conductor is

$$\varphi_0 = \frac{2\pi e^2 N}{\kappa} l^2, \tag{13}$$

where κ is the dielectric constant.

(c) The surface states charge is

$$\sigma_s = -en(\epsilon_0 - \varphi_0 - \zeta)/\epsilon_0. \tag{15}$$

(d) The total surface charge is

$$\sigma_i + \sigma_s = Nel - en(\epsilon_0 - \varphi_0 - \zeta)/\epsilon_0. \tag{25}$$

(e) The work function of the semi-conductor is:

$$\chi_2 = \varphi_s + \zeta + \frac{2\pi e^2 N}{\kappa} l^2. \tag{26}$$

Equating $\sigma_i + \sigma_s$ with $(\chi_1 - \chi_2)/4\pi ea$, we get Eq. (17) from which φ_0 or the thickness of the barrier layer, l, can be determined. With the notation:

$$l_0 = \frac{n}{N}(1 - \zeta/\epsilon_0), \tag{27a}$$

$$l_1 = \kappa\epsilon_0/2\pi e^2 n, \tag{27b}$$

$$l_2^2 = \kappa(\chi_1 - \varphi_s - \zeta)/2\pi e^2 N, \tag{27c}$$

Eq. (17) reduces to:

$$l^2/l_1 + l - l_0 = (l_2^2 - l^2)/2\kappa a, \tag{28}$$

which is a simple quadratic equation for l.

Whether the semi-conductor or the metal predominates in determining l, and thus the barrier height, depends on the relative magnitudes of the terms on the left and right sides of Eq. (28). The terms on the right side will be negligible if

$$l_1 \ll 2\kappa a \tag{29}$$

and

$$l_0 \gg l_2^2/2\kappa a. \tag{30}$$

The first condition requires that

$$n \gg \epsilon_0/2\pi e^2 a, \tag{31}$$

and the second that

$$n \gg (\epsilon_0/4\pi e^2 a)(\chi_1 - \varphi_s - \zeta)/(\epsilon_0 - \zeta). \tag{31a}$$

These conditions are essentially equivalent to that given by Eq. (23), and lead to the conclusion that $n \gg 10^{13}$ in order that the metal have little influence on the space charge region. In this case of high density of surface states, the equation for the layer thickness reduces to that for the free surface of the semi-conductor:

$$l^2/l_1 + l - l_0 = 0. \tag{32}$$

Equation (32) may, of course, be used to estimate the thickness of the space change region at the free surface of the semi-conductor, regardless of the density of surface states. The limiting case of high density corresponds to

$$l_0 \gg l_1, \tag{33a}$$

or

$$n^2 \gg N\kappa\epsilon_0^2/2\pi e^2(\epsilon_0 - \zeta). \tag{33b}$$

This is essentially the requirement, stated above Eq. (15), that n be large compared with Nl. The approximate solution of Eq. (32) is then

$$l^2 \approx l_0 l_1.$$

The product $l_0 l_1$ is independent of n, and is just the square of the thickness of the Schottky layer for a barrier height

$$\varphi_0 = \epsilon_0 - \zeta. \tag{34}$$

This is the condition that the Fermi level cross the surface near the energy corresponding to zero surface states charge.

In the limiting case of a vanishingly small density of surface states, the thickness of the barrier layer is determined by setting the right-hand side of Eq. (28) to zero. This gives

$$\varphi_0 = \chi_1 - \varphi_s - \zeta. \tag{35}$$

The height of the barrier is equal to the difference in work functions.

Equation (28) may be used for intermediate cases.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

Contact Potential Difference in Silicon Crystal Rectifiers* **

WALTER E. MEYERHOF***
University of Pennsylvania, Philadelphia, Pennsylvania

(Received February 16, 1947)

The rectifying portion of a crystal rectifier is the contact between a small point made of metal such as tungsten, and a semiconductor such as silicon or germanium containing suitable impurities. The potential energy of a conduction electron near the contact determines the rectifying action of the crystal rectifier. The most important feature of this potential energy, as far as the present paper is concerned, is the height of the potential barrier, which the electrons have to overcome when they pass from the metal to the semiconductor or vice versa. The height of the barrier is called here contact potential difference (c.p.d.), because theoretically it is equal to the difference in the work functions of the substances in contact. The c.p.d. has been measured using both n- and p-type silicon and different metallic contacts. (The c.p.d. can be obtained from the variation of the contact resistance with temperature.) The work function differences (w.f.d.) between the same substances were obtained independently by a parallel plate condenser method (Kelvin method). The results showed no correlation between the c.p.d. and the w.f.d. The c.p.d. is practically independent of the kind of metal used and also of the structure of the silicon surface. These results are in contradiction to the present theoretical model of the silicon crystal rectifier.

* Condensation of a dissertation in Physics presented to the Graduate School of the University of Pennsylvania in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Dissertation published with limited distribution under "University of Pennsylvania, BuShips Contract NObs-34144, Technical Report No. 5, August 10, 1946." Copies available from University of Pennsylvania, Department of Physics, Philadelphia 4, Pennsylvania.

** This work was done under Contracts OEMsr-388 and NObs-34144 between the Trustees of the University of Pennsylvania and the Office of Scientific Research and Development and the Navy Department, Bureau of Ships, respectively, which assume no responsibility for the accuracy of the statements contained herein.

*** Now at the University of Illinois.

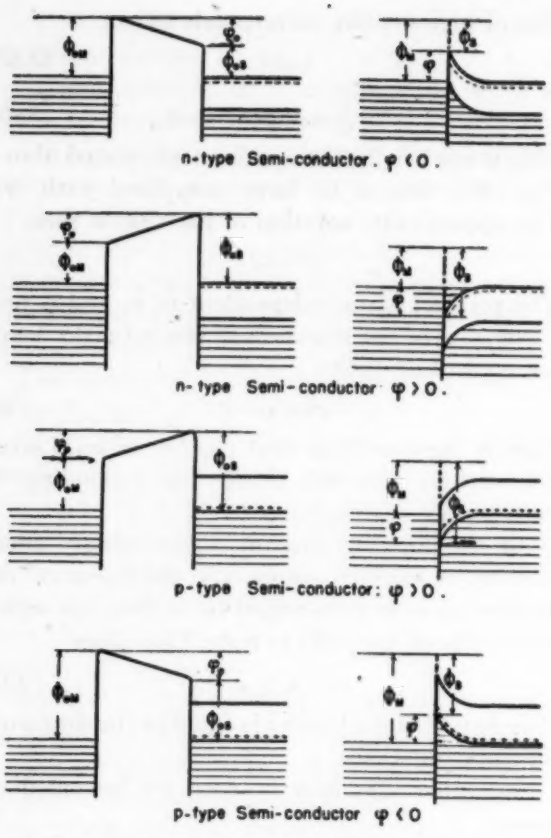


FIG. 1. Energy level diagrams for metal and semiconductor in equilibrium.

INTRODUCTION

THE essential constituents of a crystal rectifier are a metal and a semiconductor in contact. Under suitable conditions these contacts have rectifying properties.

In general one distinguishes large-area-contact rectifiers, such as copper oxide and selenium rectifiers, from point-contact rectifiers, such as silicon and germanium rectifiers, which will be described here. In both kinds of rectifier the semiconductor is in the form of a slab, making contact on both sides to a metal. Although each side will rectify in the opposite direction from the other, a net rectification is obtained in the case of the large-area-contact rectifiers by making the intrinsic rectification property of the two contacts different. In the case of point-contact rectifiers the same effect is obtained by making the area of one contact very much smaller than the area of the other.[1] Any voltage drop across the rectifier is then concentrated across the small area (high resistance) contact; in considering the theory of the point-contact rectifier one may justifiably neglect the large area contact.[1]

[1] Actually the ratio of the large to the small area must be appreciably larger than the back to front resistance ratio of the rectifier.

THEORY OF THE SEMICONDUCTOR-METAL CONTACT

When a metal and a semiconductor are brought into contact, the electron distributions adjust themselves in such a way as to make the average thermodynamic potential[2] of the conduction electrons in the solids everywhere the same. This adjustment is always accompanied by a transfer of electrons from the solid with the lower work function[3] to that with the higher work function. The transferred electrons stay in the neighborhood of the contact surface and, together with the positive charges they leave behind, form a double layer which just compensates for the original difference in work functions.

The potential (resulting from a superposition of the original lattice potential and of the potential caused by the double layer) in which the conduction electrons move has been investigated carefully by Fan.[4] In the case of silicon and germanium, though, one may make certain approximations resulting from the fact that the activation energy of the electrons (or holes) from the impurity levels into the conduction band is only of the order of a few hundredths of an electron volt[5] so that practically all the impurities are ionized at room temperature. The exact shape of the potential near the contact then depends uniquely on the distribution of impurities in the semiconductor.[6–8] For example, a uniform impurity distribution produces a parabolic potential.[6,7]

In the present considerations, it is necessary only to note that the potential jump which a conduction electron experiences in traversing the

[2] See R. H. Fowler, *Statistical Mechanics* (University Press, Cambridge, 1936), Chap. XI.

[3] Work function is defined here as the difference in potential of an electron just outside the surface of a solid and of an electron with an average thermodynamic potential energy.

[4] H. Y. Fan, Phys. Rev. 62, 388 (1942).

[5] K. Lark-Horovitz, private communication; G. L. Pearson and W. Shockley Bull. Am. Phys. Soc. [6] 21, 9 (1946); B. Serin, University of Pennsylvania dissertation, Phys. Rev. (to be published).

[6] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Clarendon Press, Oxford, 1940).

[7] H. A. Bethe, NDRC Div. 14 report, MIT Rad. Lab. 43–12, Nov. 23, 1942 (Copies obtainable from Publications Board, Department of Commerce, Washington, D. C.).

[8] B. Serin, Phys. Rev. 69, 357 (1946).

contact surface between a metal and a semiconductor is, to a first approximation, equal to the difference in work functions of the two solids. If ϕ_M is the work function of the metal and ϕ_S the work function of the semiconductor, the work function difference, φ, is defined here by

$$\varphi=\phi_S-\phi_M, \tag{1}$$

when the solids are in contact. When they are separated the same symbols with the subscript zero will be used. Figure 1 shows the various possible potential diagrams for these two situations.[9] Since the activation energies of the impurity states are only a few hundredths of an electron volt, as was mentioned above, the position of the average thermodynamic potential energy level may be taken with sufficient accuracy at the top and bottom of the forbidden zone for n- and p-type semiconductors respectively.[2,9]

The theory of the silicon or germanium point-contact rectifier was given by Bethe,[7] who pointed out that in the case of these substances the number of conduction electrons colliding with the potential barrier is appreciably larger than the number of electrons scattered by the lattice and the impurities in the barrier region. Therefore the diffusion theory, developed by Schottky,[10] Mott,[6] and others and generally applied to large-area-contact rectifiers, is not valid, but a simple diode theory will be more satisfactory for calculating the current flow across the contact. If n is the density of impurities in the semiconductor, v is the thermal velocity of the electrons perpendicular to the contact surface, e is the absolute value of the electronic charge, A is the effective contact area, k is Boltzmann's constant, T is the absolute temperature and m is the effective mass of the electron, then the current i through the contact is

$$i=j_0A\cdot\exp(-e|\varphi|/kT)\cdot[\exp eV/(kT)-1], \tag{2}$$

where

$$j_0=\tfrac{1}{2}nev=\tfrac{1}{2}ne(2kT/(\pi m))^{\frac{1}{2}}$$

for n small compared to the number of atoms per cm^3. V is the voltage drop across the contact barrier.[11] The quantity φ is called here the contact potential difference (abbreviated below by c.p.d.) and is defined by Eq. (1). It is the purpose of this paper to present some results of an experimental investigation of φ.

Actually it is necessary to take into account the effective lowering of the potential barrier at the contact because of the image force[7] and the tunneling of the electrons (or holes) through the barrier. Courant[12] has shown that under those conditions Eq. (2) must be written:

$$i=j_0A\cdot\exp(-\theta e|\varphi|/kT)\cdot[\exp(\theta eV/kT)-\exp(-(1-\theta)eV/kT)]. \tag{3}$$

θ is a numerical factor, depending on $(|\varphi|-V)$, which essentially lowers the barrier height, as shown in Fig. 2. Figure 3 is a plot of $\theta|\varphi|$ calculated by Courant for a typical contact.[13]

It is now necessary to make several remarks about the rectifier characteristic, Eq. (3). First, considering Fig. 1 it is easily seen that an n-type contact will rectify only if $\varphi<0$ and a p-type contact only if $\varphi>0$. If these conditions are not

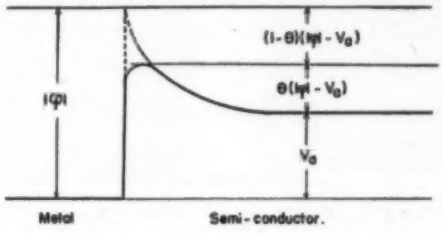


FIG. 2. Effective potential barrier for electrons or holes in a rectifying contact.

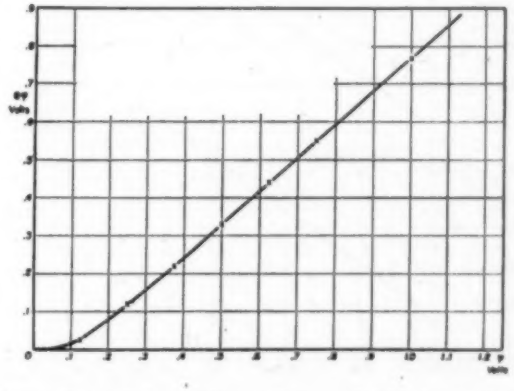


FIG. 3. Potential and observed potential for image force and tunnel effect (Courant).

[9] See also F. Seitz and S. Pasternack, NDRC Div. 14 report, University of Pennsylvania. D1–102, June 10, 1942.
[10] W. Schottky, Zeits. f. Physik 118, 539 (1942).
[11] The voltage actually applied to the rectifier is $V+iR_s$, where R_s is the spreading resistance of the "point" contact.
[12] E. Courant, NDRC Div. 14 report, Cornell University, May 17, 1943; Phys. Rev. 69, 684 (1946).
[13] $m=9.1\times10^{-28}$ g, $n=5.93\times10^{18}$ cm^{-3}, $T=290°$K, dielectric constant = 10.

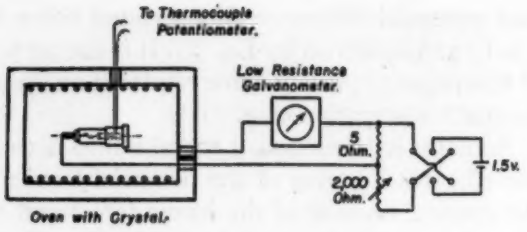


FIG. 4a. Apparatus used to measure c.p.d.

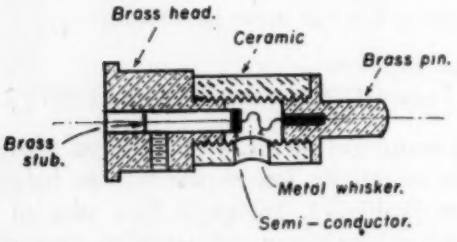


FIG. 4b. Crystal rectifier assembly.

fulfilled, there will be no effective barrier for the electrons (or holes) traversing the contact. The barrier resistance is then independent of the voltage V[11] and equal to zero. Second, in general the c.p.d. will not be uniform over the entire contact area. This so-called multi-contact effect has been examined by Sachs and others.[14] The essential result is easily seen by considering the contact as made up of two spots with different values of φ and A.[15] These spots will act like two resistances in parallel and most of the current will flow through the spot of low resistance, i.e., low $|\varphi|$, if the spot areas are not too different. This means of course that the *lowest absolute value* of the c.p.d. will determine the effective exponent $\theta e|\varphi|/(kT)$ in Eq. (3).

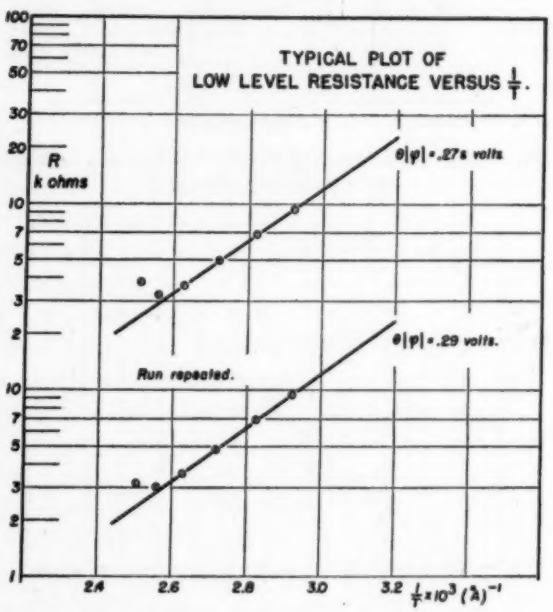


FIG. 5. Typical plot of low level resistance *versus* $1/T$. B—Be doped Si, polished, treated 2 hrs. at 975° C; 0.007" W point, ground; ½ mil. spring deflection, crystal tapped.

[14] R. G. Sachs, NDRC Div. 14 report No. 168, June 15. 1943; V. A. Johnson, R. N. Smith, and H. J. Yearian, NDRC Div. 14 report, Purdue University, August 14, 1943.

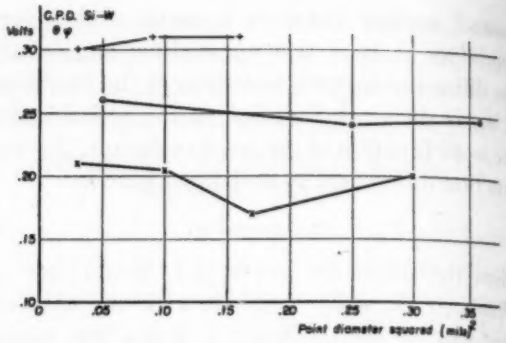


FIG. 6. Variation of observed c.p.d. Si—W as a function of W point area.

MEASUREMENT OF CONTACT POTENTIAL DIFFERENCE (c.p.d.)

In the present work it was proposed to compare the c.p.d. with the actual work function difference of a semiconductor and a metal not in contact. The c.p.d. was determined by a previously described method,[16] which consists in measuring the zero voltage resistance of the contact[17] as a function of the temperature. Equation (3) yields for this resistance:

$$R=\left(\frac{dV}{di}\right)_{V=0}=\left(\frac{kT}{j_0Ae}\right)\exp(\theta e|\varphi|/kT). \quad (4)$$

In general the linear variation of R with T can be neglected, especially since theoretically j_0 is proportional to $T^{\frac{1}{2}}$.

Figure 4a gives a diagrammatic sketch of the apparatus. The semiconductor-metal contact was mounted in a conventional temperature stable cartridge[18] shown in Fig. 4b. The oven was heated

[15] This was kindly pointed out to the author by B. Serin.
[16] W. E. Stephens, B. Serin, and W. E. Meyerhof, Phys. Rev. 69, 42, 244 (1946).
[17] The spreading resistance of the contact was subtracted from the measured resistance whenever it was not negligible compared to the barrier resistance R.
[18] See W. E. Stephens, Electronics 19, 112 (July 1946).

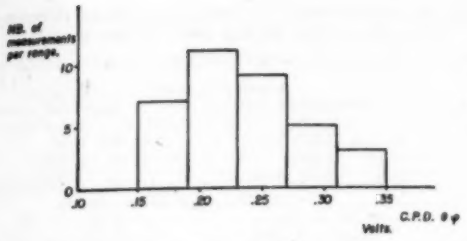


FIG. 7. Distribution curve of observed c.p.d. Si—W.

to a temperature of about 130°C in a few minutes and then allowed to cool.

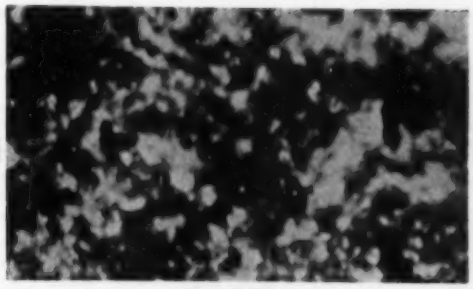
About 5 millivolts were applied to the crystal in both directions, and the average current was used to calculate the resistance of the contact, thus eliminating the effect of any thermal e.m.f. The cooling of the oven was slow enough (approx. ½ hr. from 130°C to 50°C) to permit a practically simultaneous measurement of temperature and resistance. Figure 5 shows a typical semi-log plot of R versus $1/T$. The high temperature end of the curve indicates that the semiconductor had not reached the temperature of the crystal base at the time of measurement.

RESULTS OF CONTACT POTENTIAL DIFFERENCE MEASUREMENTS

In order to study the variation of the c.p.d. with the nature of the metallic component of the contact, a particular piece of p-type silicon[19] was mounted in a cartridge and different wires were pressed against it with various pressures. The wires were previously sharpened on Arkansas stone to a point of less than 10^{-3} cm radius so that different contact pressures produced different contact areas. Figure 6 shows the variation of the c.p.d. silicon-tungsten with contact area, measured microscopically.

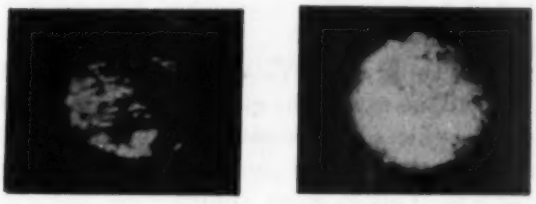
For most metals used, the c.p.d. was roughly independent of area, although there was a trend towards an increase in c.p.d. with decreasing area especially noticeable for areas less than 0.5×10^{-6} cm². This might be caused by a multi-contact effect which raises the probability of finding a larger minimum (i.e., effective) c.p.d. with small contact areas than with large contact areas.

The unavoidable electrical non-uniformity of the silicon surface itself is brought out by Fig. 7 which shows the distribution curve of approximately 50 measurements of the c.p.d. silicon-tungsten using the same piece of silicon (2 mm² area). For comparison, Fig. 8 gives an idea of the geometrical non-uniformity of different silicon surfaces and the corresponding tungsten points.



Silicon ground with No. 600 carborandum.

Silicon polished on 0000 emery paper.



Tungsten point pressed on ground silicon.

Tungsten point pressed on polished silicon.

FIG. 8. Photomicrographs of silicon surfaces and tungsten points (1000× magnification).

Table I summarizes the c.p.d. using various metals. The first column gives the measured c.p.d. $\theta\varphi$ and the second column the c.p.d. φ corrected by means of Courant's curve (Fig. 3). Each mean value of the c.p.d. (heavy type) is the average of at least 8 measurements. The extreme values of the c.p.d. measurements are also given.

Table II presents some comparison measurements on a piece of p-type and a piece of n-type

[19] 99.95 percent pure Dupont silicon, 0.002 percent boron, 0.02 percent beryllium; dry polished on 0000 emery paper; heat treated 2 hours at 975°C (see ref. 18).

TABLE I. Contact potential differences of various metals with respect to p-type silicon.

Metal	c.p.d. p-type silicon—metal Observed $\theta\varphi$ (volts)	Corrected φ (volts)
Al	0.00, **0.01**, 0.09	0.00, **0.07**, 0.21
Au	0.21, **0.23**, 0.28	0.36, **0.39**, 0.44
Cu	0.15, **0.22**, 0.36	0.29, **0.38**, 0.53
Fe	0.14, **0.17**, 0.22	0.28, **0.32**, 0.38
Mo	0.17, **0.26**, 0.39	0.32, **0.42**, 0.57
Ni	0.16, **0.31**, 0.33	0.30, **0.48**, 0.50
Ni-alloy[20]	0.21, **0.26**, 0.41	0.36, **0.42**, 0.59
Pt	0.10, **0.17**, 0.20	0.22, **0.32**, 0.35
Pt-alloy[21]	0.17, **0.19**, 0.37	0.32, **0.34**, 0.55
Ta	0.10, **0.18**, 0.26	0.22, **0.33**, 0.42
W	0.15, **0.23**, 0.32	0.29, **0.39**, 0.49

silicon treated as nearly alike as possible,[22] but differently from the piece of p-type silicon used in Table I. Only two or three measurements of the c.p.d. were made with each metal.

A comparison of the c.p.d. for the p-type silicon with that in Table I shows that the surface structure of the silicon does not seem to affect the c.p.d. (the silicon of Table II was ground and that of Table I polished—see Fig. 8 for difference in structure).

Both Table I and Table II indicate also that the c.p.d. is, within the electrical non-uniformity of the surfaces, independent of the nature of the metal for the p-type as well as for the n-type contacts. This is quite contrary to theoretical expectations considering the large variety of metals used, which have unoutgassed work functions between about 3 and 5 volts.[23] In fact the c.p.d. values for p-type silicon all lie between 0.6 and 0.0 volts. Actually the only substance giving sometimes the value $\varphi=0$ is aluminum, which has one of the lowest work functions (even unoutgassed) and which therefore should be expected to give one of the highest values of the c.p.d. according to Eq. (1).

Before considering further experiments it is of interest to present Table III which gives the change in c.p.d. when a crystal rectifier is mechanically tapped and when it is overloaded with

[20] 96.75 percent Ni, 1.00 percent Al, 2.00 percent Mo, 0.25 percent C, work hardened wire.

[21] 90 percent Pt, 10 percent Ru.

[22] For impurities in the p-type silicon see ref. 19; the impurities in the n-type piece were not known. Both pieces were ground with No. 600 carborundum and water, etched together in hot sodium hydroxide, washed and dried.

[23] See for example, A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena* (McGraw-Hill, New York, 1932).

TABLE II. Comparison of contact potential differences of p- and n-type silicon with various metals.

Metal	c.p.d. p-type silicon—metal Observed $\theta\varphi$ (volts)	Corrected φ (volts)
Pt	0.165, 0.20	0.305, 0.35
Ta	0.15, 0.20	0.29, 0.35
W	0.16, 0.18, 0.20	0.30, 0.325, 0.35
	c.p.d. n-type silicon—metal	
Pt	−0.16, −0.16	−0.30, −0.30
Ta	−0.09, −0.095	−0.22, −0.225
W	−0.07, −0.09, −0.10	−0.19, −0.22, −0.23

microwave power (i.e., "burnt out"). Only two series of measurements were made. The last two columns in Table III show the barrier resistance at room temperature and the relative contact area compared to the initial value and calculated from Eq. (4) assuming j_0 constant. It can be seen that tapping increases the c.p.d. and the resistance. Burnout decreases the c.p.d., the resistance[24] and the effective contact area.

MEASUREMENT OF WORK FUNCTION DIFFERENCE (w.f.d.)

Since the above mentioned experiments showed no correlation of the c.p.d. between silicon and various metals, and the work functions of the metals, it was thought advisable to measure the work function difference (abbreviated below by w.f.d.) directly, under various conditions. Although these measurements eventually did not suggest any new phenomena which would make the c.p.d. roughly independent of the silicon surface and the nature of the metal, they demonstrate very clearly the disagreement of the c.p.d. measurements with the presently assumed model of the contact.

TABLE III. Change of contact potential difference of a p-type silicon-tungsten rectifier with tapping and power overloading.

Treatment crystal rectifier	c.p.d. p-type silicon—tungsten Observed $\theta\varphi$ (volts)	Corrected φ (volts)	Resistance R (k-ohm)	Relative contact area A/A_{100} (%)
Initial values	0.21, 0.17	0.36, 0.32	4.8, 2.4	100, 100
After tapping	0.28, 0.20	0.44, 0.35	37, 17	200, 47
After 7 watt RF peak power[25]	0.18, 0.14	0.33, 0.28	6.4, 3.8	23, 28
After 70 watt RF peak power	0.08, 0.10	0.20, 0.23	0.4, 0.9	6, 15

[24] For a more detailed study of resistance as a function of burnout see A. W. Lawson *et al.*, NDRC Div. 14 report No. 113, November 1, 1942.

[25] 10 cm microwave microsecond pulses were used in this experiment; the power given is available peak power.

TABLE IV. Variation of work function difference between p-type silicon and various abraded metals.

Metal	w.f.d. p-type silicon—metal 1 min. after abrasion of metal φ_0 (volts)	50 hrs. after abrasion of metal φ_0 (volts)	Change in ϕ_{0M} in 50 hrs. in air (volts)	Change in ϕ_{0M} caused by adsorbed oxygen (Dubois) (volts)	Change in ϕ_{0M} caused by adsorbed water vapor (Dubois) (volts)
Al	0.75	0.21			
Al[28]	0.69	0.26			
Au	−0.31	−0.45	0.14	0.14	−0.14, −0.21
Cu	−0.37	−0.42	0.05	0.15, 0.16	−0.03, −0.06
Cu*	−0.34	−0.40	0.06		
Fe	−0.18	−0.41	0.23	0.22, 0.34	−0.56, −0.56
Mo	−0.17	−0.42	0.25	0.18	−0.16, −0.18
	−0.21	−0.38	0.17		
Ni	−0.09	−0.18	0.11	0.15	−0.16
	−0.08	−0.12	0.04		
Ta	0.13	−0.04			
Pt	−0.60	−0.61	0.01	0.04	−0.02
W	−0.19	−0.33			

The method of measurement of the w.f.d. was essentially a parallel plate condenser method (Kelvin method). The two substances under investigation form the two plates of a variable condenser. The field established by the w.f.d. (see Fig. 1) is brought to zero by an externally applied voltage (equal to the w.f.d.). The details of the method are described elsewhere.[27] In opposition to the c.p.d. measurement, this method gives practically the *average* w.f.d. between two surfaces.

The surfaces under study were usually abraded with 000 emery cloth in order to produce a surface treatment similar to grinding the metal points in the c.p.d. experiments. Also an apparatus was constructed which permitted this abrasion in a vacuum of about 5×10^{-5} mm Hg.

TABLE V. Variation of work functions of tungsten and silicon after abrasion.

Gas	Initial impurities	Increase in work function in 10 min in gas after abrasion in vacuum: Tungsten (volts)	Silicon (volts)
Air	—	0.33, 0.41	0.10, 0.12 p-type 0.22, 0.26 n-type
Oxygen	—	0.42, 0.46	0.11, 0.15 p-type
Oxygen and Water vapor	—	0.46, 0.50	0.16, 0.18 "
Nitrogen	0.1% O_2, 0.4% A	0.14, 0.16	0.10, 0.12 "
Hydrogen	0.1% O_2	0.08, 0.12	0.08, 0.08 "
Helium	1.6% N_2, 0.2% hydrocarbons	0.06, 0.08	0.04, 0.06 "
Argon	0.5% N_2, 0.001% O_2, H_2, H_2O	0.04, 0.06	0.03, 0.05 "

[28] All metals guaranteed pure except the starred metals which are commercial stock. The purity of the metal does not influence the work function to any appreciable extent in these experiments.

[27] W. E. Meyerhof and P. H. Miller, Jr., Rev. Sci. Inst. 17, 15 (1946).

TABLE VI. Data for the determination of the absolute work function of a silicon surface.

Metal	Oxide coating in air	Work function of oxide (Lammermann and Lange)	w.f.d. silicon—metal (Table IV) φ_0 (volts)	Work function of silicon surface ϕ_{0S} (volts)
Al	Al_2O_3	3.9[28]	0.21, 0.26	4.11, 4.16
Cu	Cu_2O	4.47	−0.42, −0.40	4.05, 4.07
Fe	FeO or Fe_3O_4	4.54 or 4.58	−0.41	4.13 or 4.17
Pb[29]	PbO	4.06	0.13	4.19

RESULTS OF WORK FUNCTION DIFFERENCE MEASUREMENTS

It is recalled that the w.f.d., φ_0, is defined by an equation similar to Eq. (1):

$$\varphi_0=\phi_{0S}-\phi_{0M},$$

where ϕ_{0S} is the work function of the silicon and ϕ_{0M} is the work function of the metal, as measured in the present experiment.

Table IV summarizes the experiments in air (on the same metals as used in Table I). A piece of polished p-type silicon[19] was taken as standard surface. Each metal was abraded with 000 emery cloth and the w.f.d. measured approximately 1 min. and 50 hrs. after abrasion. The results on Mo and Ni in the above table give an idea of the reproducibility of the results. The fourth column in Table IV gives the change in the metal work function during 50 hours exposure in air. For comparison the last two columns show some results of Dubois[30] on the influence of oxygen and water vapor on outgassed work functions. It can be inferred from these results that the increase in work function of the metals after abrasion in air is mostly caused by (pure or activated) adsorption of oxygen and that the humidity in the air does not play any significant role.

This is also brought out by Table V which gives the change in work function of tungsten and silicon (each measured against a standard surface) abraded in a vacuum of about 5×10^{-5} mm Hg and then subjected to various gases at atmospheric pressure for 10 minutes. The numbers

[28] Compare with 3.95 volts obtained by H. J. Spanner, Ann. d. Physik 75, 608 (1925).

[29] Not shown in Table IV.

[30] E. Dubois, Ann. de physique 14, 627 (1930).

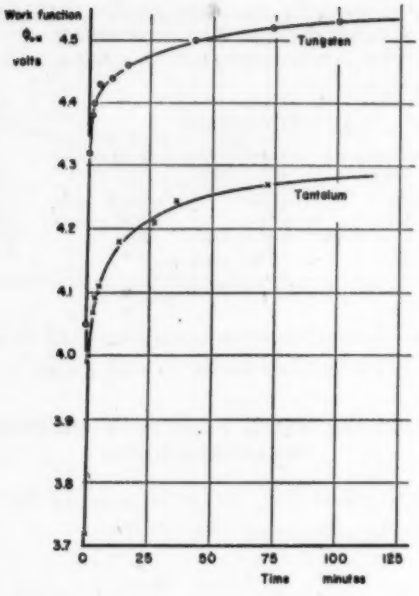


FIG. 9. Work function of tungsten and tantalum after abrasion in partial vacuum.

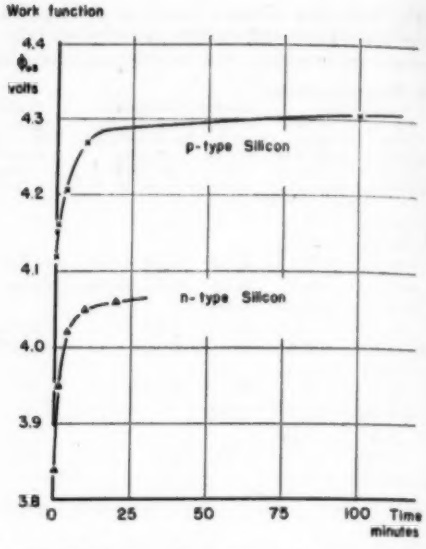


FIG. 10. Work function of p- and n-type silicon, after abrasion in partial vacuum.

given represent averages of two experiments. The gases in which oxygen was an impurity were passed over red hot copper, so that the percentage oxygen present was certainly less than shown. Table V would indicate then that nitrogen and hydrogen can also affect the work function of solids.[31]

It is known that most metals oxidize in air. Lämmermann and Lange[32] have measured the work functions of several oxides and their values allow a calibration of the standard silicon surface used in Table IV, if we assume that the metals in the present experiment were coated with the oxides shown below, 50 hrs. after abrasion in air. From Table VI a work function between 4.1 and 4.2 volts may be obtained for the standard silicon surface.

Figures 9 and 10 show the variation of work function of tungsten, tantalum, n- and p-type silicon after abrasion in a vacuum and then subjecting the surfaces to air at atmospheric pressure within ½ min. The measured w.f.d. was calibrated by using the standard silicon surface and calling its work function 4.2 volts. The inaccuracy of this value will evidently not affect the relative scales of Figs. 9 and 10. These figures show that for each solid the work function increases on exposure to air.

DISCUSSION OF EXPERIMENTAL RESULTS

Several comments on these measurements are now in order. First, assuming the effect of abrasion to be the same on the surfaces of both n- and p-type silicon, the difference in work functions of the two solids should be equal to the width of the forbidden region in silicon (1.12 volts[33]). Figure 10 shows that it is only about 0.3 volts. Second, since the work functions of the solids examined change very markedly on exposure to air, one might assume that on forming a contact between a metal and silicon, there occurs a change in the oxide layers in such a way as to make the results of Tables I and II consistent with those of Table IV and Figs. 9 and 10.

Holm[34] has found that in the case of metal-metal contacts the plastic deformation of the metal parts actually in contact causes a breaking of the oxidation layers. Let us assume a similar simple mechanism in the formation of crystal rectifier contacts. Since the silicon is too hard to

[31] Since nitrogen is generally an impurity in argon this might invalidate some of the conclusions which O. Klein and E. Lange, Zeits. f. Elektrochemie **44**, 542 (1938), drew from a change of the work function with time in an atmosphere of argon.

[32] Values given by O. Klein and E. Lange (reference 31).

[33] F. Seitz, NDRC Div. 14 report No. 113, November 1, 1942.

[34] R. Holm, *Die technische Physik der elektrischen Kontakte* (Julius Springer, Berlin, 1941; reproduced by J. W. Edwards, Ann Arbor, 1944), pp. 63–139.

be deformed, we may assume that its oxidation layer is conserved during contact formation, so that a crystal rectifier contact is perhaps roughly a silicon "oxide"-metal contact. Table VII compares the w.f.d. (assuming this kind of contact and assuming a silicon "oxide"-metal "oxide" contact) with the c.p.d. (Table II) for the same metals. (The work functions of the p- and n-type silicon were measured after treating the surfaces similarly to those of Table II[22] and were found to be 4.13 and 3.95 volts respectively.) It is easily seen that the w.f.d. measurements cannot be made consistent with the c.p.d. measurements. In this connection it should be noted that the multi-contact theory requires for a p-type contact $\varphi=0$, if $\varphi_0 \leqslant 0$, and for an n-type contact $\varphi=0$, if $\varphi_0 \geqslant 0$.

The general nature of the results is in essential disagreement with some experiments on large-area contact rectifiers performed by Joffe.[35] Joffe finds that there is agreement between the sign of the w.f.d. and the finite or zero value of the corresponding c.p.d. It might not be justified, though, to compare those results (on semiconducting oxides mostly) with the present experiments (on silicon).

CONCLUSION

Assuming the correctness of the experimental results, it seems desirable to reinvestigate the physical picture of the silicon-metal contact. Essentially it is necessary to construct a model which makes the c.p.d. roughly independent of the surface structure of the silicon and roughly independent of the work function of the metal. It is possible to construct such a model by assuming the surface of the silicon to be practically

[35] A. V. Joffe, J. Phys. USSR **10**, 49 (1946).

TABLE VII. Comparison of values for contact potential difference and work function difference.

Nature of contact	c.p.d. (Table II) φ (volts)	w.f.d. for particular model of contact: Silicon "oxide"-metal φ_0 (volts)	w.f.d. for particular model of contact: Silicon "oxide"-metal "oxide" φ_0 (volts)
p-Si—Pt	0.3	—	−0.7
p-Si—Ta	0.3	0.4	−0.1
p-Si—W	0.3	0.1	−0.4
n-Si—Pt	−0.3	—	−0.9
n-Si—Ta	−0.2	0.2	−0.3
n-Si—W	−0.2	−0.1	−0.6

"metallic" [36] so that the potential barrier would be located inside the silicon surface. Any metal making contact to such a surface would introduce no further discontinuities in the potential, just as in the case of metal-metal contacts, so that the c.p.d. would be independent of the work function of the metal.

The thickness of the "metallic" layer would not have to be more than a few atomic distances in order to satisfy the electronic conditions, so that this model essentially assumes the existence of particular surface states[37] for the conduction electrons. Calculations in this general direction might throw further light on the mechanism of silicon point-contact rectifiers.[38]

ACKNOWLEDGMENTS

The author wishes to thank Dr. W. E. Stephens for his constant advice and guidance in this research and Dr. L. I. Schiff for his many valuable suggestions.

[36] This was kindly pointed out to the author by G. H. Wannier.

[37] See for example, F. Seitz, *The Theory of Solids*, (McGraw-Hill Book Company, New York, 1940), sec. 70.

[38] Note added in proof: Since this paper was sent in for publication, J. Bardeen has kindly informed the author of his calculations on surface states, which appear in Phys. Rev. **71**, 717 (1947).

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is, for the issue of the 1st of the month, the 8th of the preceding month and for the issue of the 15th, the 23rd of the preceding month. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length.

The Scattering of Neutrons of Energy Between 12 Mev and 13 Mev by Protons

J. S. LAUGHLIN AND P. G. KRUGER
University of Illinois, Urbana, Illinois
April 7, 1947

THE angular distribution of the recoil protons in the neutron-proton scattering process has been investigated in a high pressure (25 atmospheres) cloud chamber. The recoil protons were produced in methane gas by a well-collimated homogeneous beam of neutrons one-half inch in diameter taken in the forward direction from the $D(d, n)He^3$ reaction. A thin (0.5-Mev) deuterium gas target was employed and a magnetic field deflected the deuteron beam from the forward direction after passing through the target. This prevented neutrons from sources other than the deuterium target from entering the cloud chamber. The neutron spectrum of this arrangement was examined with the cloud chamber and showed a strong peak between 12.0 Mev and 13.0 Mev, superimposed on a low intensity continuous spectrum originating from the thin-foil windows on the entrance and exit of the target.

A total of 2000 recoil protons have been measured, of which 1636 were produced by neutrons between 12-Mev and 13-Mev energy and satisfied selection criteria. The data are given in Table I.

The aximuthal correction factors given were determined from the experimental data and are in good agreement with the correction factors calculated from the geometry and the illuminated height of the chamber. For each scattering angle interval the data were plotted in 20° azimuthal angular intervals. The shape of this distribution was in good agreement with the calculated shape based on purely geometrical considerations. The number of recoil protons scattered into those azimuthal angle intervals in which all tracks were observable determined the number which would have been measured if all the tracks in all the azimuthal angle intervals could have been measured. The ratio of this number to the actual observed number yielded the azimuthal correction factors.

TABLE I. Experimental data for recoil protons produced by neutrons of energy between 12 Mev and 13 Mev.

Angular interval c.m. system	No. of recoil protons	Azimuthal correction factors	Length correction factors	Solid angle interval $\times 1/2\pi$	No. of recoil protons per unit solid angle $\times 2\pi/10$
0–21	96	1.12	2.78	0.066	452 ± 10.5%
21–41	174	1.94	2.32	0.179	438 ± 9.3%
41–61	280	2.25	1.82	0.270	425 ± 7.0%
61–81	447	2.07	1.39	0.320	391 ± 5.7%
81–101	639	1.78	1.18	0.347	387 ± 5.1%

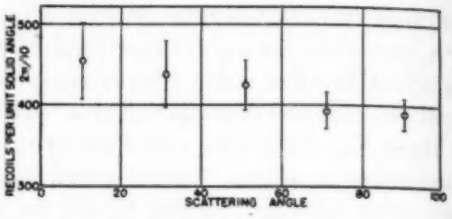


FIG. 1. Angular distribution of recoil protons per unit solid angle in the center of mass system.

Since the neutron beam is collimated, all recoil protons originate from a ½-inch diameter cylinder through the center of the chamber, and since the light beam height is known, a calculation of the expected azimuthal correction factor is quite straightforward.

A length correction factor is necessary since for each different scattering angle interval the recoil protons have a different energy and length. Since these lengths and the chamber geometry were well known, these factors could be calculated from simple geometrical considerations.

The graph in Fig. 1 shows the number of recoil protons per unit solid angle per scattering angle interval in the center of mass system plotted *versus* the scattering angle in the center of mass system. Probable statistical errors are indicated in each interval. A ratio of the differential cross sections for neutron scattering in the backward direction to scattering in the perpendicular direction in the center of mass system of either unity (spherically symmetric scattering) or slightly greater than unity can be consistent with these data. This possible small asymmetry is in the opposite direction to that found by Amaldi[1] and co-workers and by Champion and Powell.[2] Later results by Powell[3] apparently show spherically symmetric scattering. A complete discussion of these data and the analysis will be published later.

[1] E. Amaldi, D. Bocciarelli, B. Ferretti, and G. C. Trabacchi, *Naturwiss.* **30**, 582 (1942).
[2] F. C. Champion and C. F. Powell, Proc. Roy. Soc. **A183**, 64 (1944).
[3] Reference 38, G. Wentzel, Rev. Mod. Phys. **19**, 10 (1947).

Erratum: The *K* X-Ray Absorption Edge of Silicon

[Phys. Rev. **71**, 406 (1947)]
VOLA P. BARTON, *Goucher College, Baltimore, Maryland*
AND
GEORGE A. LINDSAY, *University of Michigan, Ann Arbor, Michigan*

IT was erroneously stated in the above paper that in Table I the wave-lengths were corrected only for normal refraction. Actually the terms for the anomalous effect had also been applied for quartz and the four micas, which are the only ones needing such a correction. The anomalous terms increased the observed wave-lengths by 0.6 x.u. for quartz and by 3.0 x.u. for the micas.

Čerenkov Radiation Counter*

R. H. DICKE
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
April 23, 1947

GETTING[1] has proposed a high speed fast-particle counter using, to actuate a photo-multiplier tube, Čerenkov radiation[2] produced as a charged particle traverses a plastic rod. The advantages of such a system for fast counting, timing, and velocity measurement are evident. Furry has recently proposed a measurement of meson mass using a modification of the velocity measurement feature.[3]

A device similar to Getting's Fig. 3 was tried at Princeton last summer. The apparatus consisted of a Geiger counter, Lucite flared rod, lens system, and photo-multiplier, all in a vertical line. A fast charged particle would actuate the Geiger counter, then produce in the Lucite Čerenkov radiation which would in turn be detected by the photo-multiplier and associated amplifiers. The apparatus was run for 4500 hours between August 1, 1946, and October 18, 1946, with amplifier discrimination set at various levels; coincidence counts definitely attributable to Čerenkov radiation were not obtained.

Recently another attempt to demonstrate operation of the device was made using electrons generated in lead by x-rays from a 20-Mev betatron. In Fig. 1 is shown the arrangement of components. The x-rays strike a ¼-inch piece of lead; electrons and positrons generated therein pass down a collimating hole in a lead plug and then into the Lucite rod. Čerenkov radiation from the rod is focused onto the 1P28 photo-multiplier. The latter operates at liquid nitrogen temperature in an evacuated envelope not shown. Photo-multiplier pulses are fed from a cathode follower into a long transmission line, thence into a Los Alamos Model 501 amplifier. Pulses from the amplifier are fed into a scaler with a calibrated pulse height discriminator. The betatron, run at reduced intensity, was monitored by a Geiger counter.

Figure 2 shows the result of one set of measurements. Curve *A* represents the number of pulses greater than a certain height plotted against this height. Each point was taken for a definite Geiger-monitored x-ray flux. Curve *B* is obtained from *A* by subtracting dark pulses observed with the betatron off. Curve *C* is obtained from *B* by subtracting the x-ray background counts measured by interposing a thin piece of blackened cardboard between lens and photo-tube (Fig. 1). Curve *C* thus represents the distribution of photon pulses. As a check on "pile up" a run was made with betatron intensity reduced by a factor of 2.5; the resulting photon distribution is represented by the circles in Fig. 2.

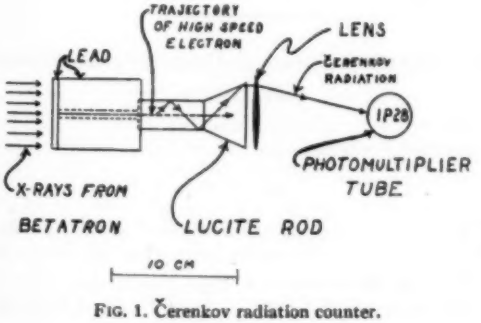


FIG. 1. Čerenkov radiation counter.

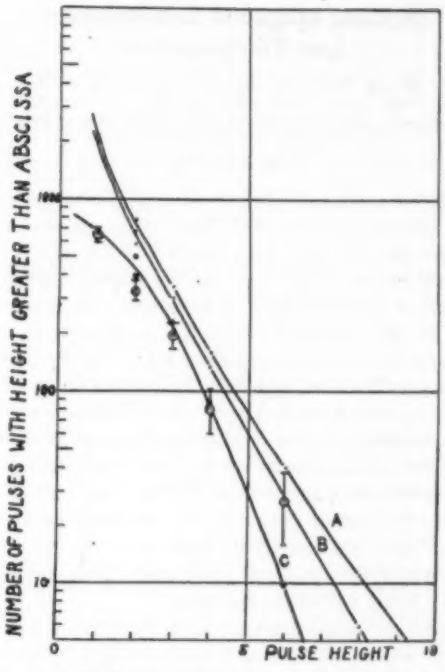


FIG. 2. *A*—total pulses. *B*—total pulses minus dark pulses. *C*—light pulses (total minus dark minus x-ray background). Circles—light pulses from reduced intensity run.

Curve *C* is in agreement with the energy distribution of electrons produced in lead by bremsstrahlung from 20-Mev electrons. Comparison of curve *C* with the distribution of pulses from single photons (unfortunately not taken at the same time) indicates probably no more than 5 photo-electrons in the largest pulses. This is about ¼ of the computed[3] maximum (for photoelectric efficiency see RCA Tube Manual).

In conclusion it can be stated that photon pulses are obtained from the counter shown in Fig. 1. The data are in reasonable agreement with the hypothesis that the photon counts are caused by Čerenkov radiation. Considering the geometry of the counter it seems unlikely that fluorescence is an important source of photons. However, fluorescence cannot be definitely ruled out as a contributing factor.

The author wishes to express his appreciation to Colonel Carl Dutton and Miss Kathleen George of the Picatinny Arsenal for their hospitality and helpful assistance in making available the Kerst betatron at Picatinny. He is indebted to Dr. W. C. Elmore for aid with electronic problems, and to Dr. D. W. Kerst for advice and many helpful suggestions.

* This work was supported in part by Navy Contract N6OR1-105 Task 1.

[1] I. A. Getting, Phys. Rev. **71**, 123 (1947).

[2] P. A. Čerenkov, Comptes Rendus U.S.S.R. **8**, 451 (1934). I. Frank and I. Tamm, Comptes Rendus U.S.S.R. **14**, 109 (1937). G. B. Collins and V. G. Reiling, Phys. Rev. **54**, 499 (1938).

[3] W. H. Furry, Bull. Am. Phys. Soc. **22**, No. 2, p. 27 (1947), to be published in *The Physical Review*.

Nuclear Magnetic Resonances at Low Temperatures*

F. Bitter, N. L. Alpert, H. L. Poss, C. G. Lehr, and S. T. Lin**
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
April 14, 1947

WE have recently completed a preliminary set of experiments on nuclear magnetic resonances at temperatures down to 4°K, using the 9½-inch gap of the M.I.T. cyclotron magnet which was recently made available to us for a period of several days. Following is a summary of the results of these experiments and of a series of room temperature experiments done with a small electromagnet.

Proton or fluorine resonances have been observed in all substances investigated. Among these have been: benzene, ether, paraffin, lithium hydroxide, lithium fluoride, and hydrofluoric acid at room temperature; glycerine down to 195°K (solid CO_2); ammonium chloride, water and ice both as pure H_2O and in the form of a mixture containing 95 percent D_2O+5 percent H_2O down to 77°K (liquid nitrogen); liquid hydrogen at its boiling point; and solid methane down to liquid helium temperature.

In general, broad resonance lines, many gauss wide, were found in solids, with the exception of CH_4, while lines of the order of one gauss and less in width were observed in liquids. In reference to the distilled water and heavy water solutions, the D_2O was found to have no appreciable effect on the width of the resonance line in both the liquid and frozen states. The glycerine resonance line broadened from a width of the order of one gauss to a width of about 16 gauss as the temperature was reduced to 195°K. The LiF gave the broadest resonance observed, greater than 25 gauss wide at room temperature. The resonance line in liquid hydrogen was slightly broader than for protons in water at room temperature. In most cases, narrow lines were observed on an oscilloscope. Because of the lower amplitude of broad resonances, these were investigated on the output meter of a narrow-band 30-cycle amplifier in order to obtain a better signal to noise ratio.

Our observations of narrow lines in liquids and broad lines in solids are in agreement with the theory of line widths outlined by Bloembergen, Pound, and Purcell.[1] In this theory the rotational states in liquids are the key to narrow lines. In solids the internuclear fields serve to broaden the observed resonance line. However, when rotations occur with frequencies greater than the Larmor frequency involved (about 30 Mc for these experiments), the internuclear fields practically average out to zero during one Larmor cycle, thus resulting in a narrow line.

Since rotational states generally lead to narrow lines, it was decided to study the possibility of rotational states above the λ-points of ammonium chloride and solid methane, which have specific heat anomalies at −30°C and 20°K, respectively. Although some slight broadening of the proton line occurred in cooling ammonium chloride from room temperature to 77°K, it was a consistently broad line, such as is characteristic of normal solids. This indicates that the transition across the λ-point is due to an order-disorder phenomenon, in agreement with conclusions drawn by Lawson[2] from an analysis of thermodynamic data. In contrast to this are the observations of proton resonances in solid methane. In a liquid helium bath a broad line, about 12 gauss wide, was found. When the sample was allowed to warm up through the λ-point, the broad line changed to a narrow line, presumably in the vicinity of the λ-point. At liquid hydrogen and nitrogen temperatures, narrow lines, somewhat less than 1 gauss wide, were observed in solid methane.

It should be noted that those molecules with H atoms showing λ-point transitions are among the easiest to study by this method but present considerable difficulty in studies by means of x-rays. More detailed investigations of rotational states in solids are now being undertaken.

Larger signals were observed at low temperatures, approximately following the inverse temperature variation in Curie's formula for the equilibrium magnetization. This is in agreement with results obtained at hydrogen temperatures by Rollin and Hatton[3] and indicates that, in the substances examined, relaxation times do not undergo drastic variations at low temperatures. In the case of resonances in liquid hydrogen, signal-to-noise ratios of from 300 to over 1000 were obtained on an oscilloscope without the use of a narrow-band amplifier. These results indicate that low temperatures may be useful in the study of small samples or of substances with low concentrations of the nucleus under observation, particularly if narrow lines can be retained as in methane. However, in most substances, broad resonance lines are to be expected, and this will tend to reduce the gain caused by the low temperatures. If the structure of resonance lines is to be studied, or if the natural width of narrow lines is to be measured, stringent requirements must be placed on the uniformity of the magnetic field over the sample. Otherwise the line width and shape will be determined by the magnet and not by the sample.

We are indebted to Professor E. M. Purcell of Harvard for his suggestions in reference to the r.f. circuit which we are using. We also wish to thank many members of the M.I.T. staff, particularly Professor C. F. Squire for his assistance in the low temperature experiments, and Dr. E. T. Clarke, through whose cooperation we were able to use the cyclotron magnet.

* This research has been supported in part through the Joint Service Contract, No. W-36-039 PC-32037.
** On leave from the Institute of Physics, Academia Sinica, China.
[1] N. Bloembergen, R. V. Pound, and E. M. Purcell, Phys. Rev. 71, 466(A) (1947).
[2] Rollin and Hatton, Nature 156, 201 (1947).
[3] Andrew W. Lawson, Phys. Rev. 57, 417 (1940).

The Unboundedness of Quantum-Mechanical Matrices

Aurel Wintner
Department of Mathematics, The Johns Hopkins University, Baltimore, Maryland
April 9, 1947

HILBERT'S definition of a *bounded* infinite matrix A is equivalent to the requirement that the linear substitution $y=Ax$ should define a vector $y=(y_1, y_2, \cdots)$ of finite length whenever $x=(x_1, x_2, \cdots)$ is a vector of finite length

(i.e., a point of Hilbert's space). Direct inspections show that all those quantum-mechanical matrices Q, P which have been calculated explicitly (linear oscillator, hydrogen atom, etc.) fail to be bounded in Hilbert's sense.

It is natural to expect that this failure is independent of the explicit choice of the particular Hamiltonians H in question. In fact, the unboundedness of Q, P is indicated by heuristic considerations concerning the location of possible energy levels. However, the literature consulted does not contain a general proof. The purpose of this note is the construction of such a proof.

If Planck's constant is expressed in appropriate units, Heisenberg's relation can be written in the form

$$QP-PQ=\tfrac{1}{i}E, \tag{1}$$

where E denotes the unit matrix and Q, P are supposed to be Hermitian. Without any reference to a Hamiltonian H, it will be shown that these assumptions prevent the boundedness of Q *and* P.

Let A^* denote the transposed matrix of the complex-conjugate of an arbitrary matrix A. Thus A is Hermitian if and only if $A^*=A$. Since Q and P are Hermitian, it follows that A^* is $Q-iP$, if A denotes $Q+iP$. Hence, Eq. (1) can be written in the form

$$A^*A-AA^*=E. \tag{2}$$

Furthermore, if P and Q are bounded, then A is bounded. It will be shown that these assumptions lead to a contradiction.

By the spectrum of a bounded matrix A is meant the set of those real or complex values λ for which the bounded matrix $\lambda E-A$ fails to have a (unique) bounded reciprocal $(\lambda E-A)^{-1}$. It is known that the spectrum of every bounded matrix is contained in a sufficiently large circle about the origin of the λ-plane, and that every bounded matrix *has* a spectrum[1] (i.e., that the spectrum contains at least one λ-value). Because of the first of these two properties, $A-\lambda E$ must have a bounded reciprocal matrix if the absolute value of λ is large enough. On the other hand, it is clear that the expression to the left side of Eq. (2) remains unaltered if A is replaced by $A-\lambda E$, where λ is any real value. Hence, if $A-\lambda E$, where λ is real and large enough, is denoted by A, there results a bounded matrix A satisfying Eq. (2) and possessing a bounded reciprocal A^{-1}. But Eq. (2) can then be written in the form

$$B-ABA^{-1}=E, \tag{3}$$

where $B=A^*A$. In fact, the product of bounded matrices is bounded, and their multiplication is associative. In particular, B is bounded.

It is clear from the above definition of a spectrum that if A, A^{-1} and B are bounded, then B and its transform ABA^{-1} have the same spectrum. Since Eq. (3) means that ABA^{-1} is identical with $B-E$, it follows that B and $B-E$ have the same spectrum. Finally, the definition of a spectrum also shows that the spectrum of $B-E$ results from that of B by a rigid motion, representing a translation (of unit span) of the λ-plan into itself.

Accordingly, the spectrum of B remains invariant under a certain translation. Consequently, either the spectrum of B contains every point of a line in the λ-plane or the spectrum of B has no point at all. However, as mentioned above,[1] both of these possibilities are ruled out by the boundedness of B. This contradiction proves the unboundedness of $B=A^*A$, and therefore that of $A=Q+iP$.

[1] A. Wintner, *Spektraltheorie der Unendlichen Matrizen* (Hirzel, Leipzig, 1929) pp. 143–44.

On the Dependence of Nuclear Radius on the Mass Number

E. Amaldi and B. N. Cacciapuoti
Istituto di Fisica della Università, Centro di Fisica Nucleare del C.N.R., Rome, Italy
April 17, 1947

In an earlier paper[1] we reported some measurements on nuclear radius for 19 elements with Z between 4 and 83, and sought a linear representation of our results by an expression of the form

$$R=b+r_0A^{\frac{1}{3}}. \tag{1}$$

By graphical interpolation, in which we intentionally enhanced the importance of results corresponding to elements with high atomic number, we obtained for b and r_0 the following values:

$$b=1.5 \text{ to } 2.0\times10^{-13} \text{ cm}, \quad r_0=1.3 \text{ to } 1.0\times10^{-13} \text{ cm}. \tag{2}$$

The quantity b represents the surface effect of the nucleus, that is, the contribution of nuclear force range of superficial particles to the value of R; in fact our measurements refer to a nuclear radius R defined as the distance from the center of a nucleus at which a fast neutron interacts with the nucleus. Other definitions may be given; for instance,[2] when the binding energies of contiguous and specular isobar pairs are compared, the nuclear radius can be defined as the radius of the proton-distribution sphere. In all cases in which a comparison was possible, the values obtained using this second definition were found to be smaller than ours by about 1×10^{-13} cm.

In our former paper[1] it was mentioned that, notwithstanding the accuracy of the values which we found for R, it is not possible to determine the values of b and r_0 with great precision, owing to the appreciable deviations of the individual points from the linear representation of Eq. (1).

Therefore we thought it worth while to treat the question again by statistical methods in order to test the correctness of fit of a theoretical curve to a set of experimental values, although such procedures are generally open to some criticism.

The second column of Table I gives the values of $R_i+\delta R_i$ for 17 elements. The results obtained for C and Au have been omitted; the geometrical conditions for the C measurements were not quite the same as for the other elements, and the results obtained with Au were much less accurate than for other elements.

The third column gives the ratio of the statistical error to the standard deviation; the values δR_i of the second column are in all cases the greater of the two errors. A detailed account of the calculation of these two errors was given in reference 1; we shall only stress the fact that in

TABLE I.

Element	$(R_i \pm \delta R_i) \cdot 10^{-13}$ cm	Statistical error Standard deviation
4 Be	3.22 ±0.074	1.5
5 B	4.30 ±0.24	—
12 Mg	5.40 ±0.13	0.55
13 Al	5.54 ±0.11	0.63
16 S	5.02 ±0.14	2
26 Fe	6.63 ±0.06	1.2
28 Ni	6.47 ±0.074	1.5
29 Cu	6.75 ±0.15	0.3
30 Zn	6.96 ±0.17	3
34 Se	7.31 ±0.20	0.55
47 Ag	7.81 ±1.09	1.3
48 Cd	8.23 ±0.07	2.3
50 Sn	8.50 ±0.085	4
51 Sb	8.33 ±0.096	10
80 Hg	9.50 ±0.16	0.58
82 Pb	9.00 ±0.071	0.87
83 Bi	9.08 ±0.097	3.7

about half of the cases we found that this ratio was smaller, and in the other half was greater than unity, so we believe that the main error is purely statistical. This means that the probability for the real value to deviate by an amount ΔR_i from the observed value may be represented by means of a normal distribution (Gaussian law).

Under this reasonable assumption, we may calculate the probability that a given distribution will deviate from the interpolated straight line as much, or more, than our observed points.

Such a probability is obviously given by

$$P = \int_{(\chi^2)_0}^{\infty} e^{-\frac{1}{2}\chi^2} \chi^{n-1} d\chi \Big/ \int_0^{\infty} e^{-\frac{1}{2}\chi^2} \chi^{n-1} d\chi, \qquad (3)$$

where

$$\chi^2 = \sum_1^n (\Delta R_i / \delta R_i)^2 \qquad (4)$$

and n is the number of measured points used for comparison with the interpolated straight line of Eq. (1); in our case $n=17$. The deviations of the observed points from the interpolated line are indicated with ΔR_i and the corresponding experimental errors are δR_i (see the second column of Table I).

The function defined by (3) is formally identical to the function P introduced by Pearson for his "χ^2 test," although in our case the definition of χ^2 is slightly different from his. In order to calculate P, the most convenient linear representation (1) is the one that makes χ^2 a minimum. According to the definition (4) of χ^2, this obviously means that the constants b and r_0 of (1) must be determined by the least-squares method, considering the different precision δR_i of the single points.

In this way we get

$$b = (0.696 \pm 0.082) \times 10^{-13} \text{ cm},$$
$$r_0 = (1.52 \pm 0.02) \times 10^{-13} \text{ cm}, \qquad (5)$$

that is, two values quite different from those obtained with the graphical interpolation (2), in which we intentionally over-rated the points corresponding to heavy elements.

Applying definition (4), with the experimental values of Table I, and through the representations (1) and (5), we obtain for χ^2 the value $\chi^2 \leqq 270$ which corresponds, for $n=17$, to a very small value of P defined by (3). From Pearson's tables,[3] it is seen that, for $n=17$, P is smaller than 10^{-6}, even for $\chi^2=70$.

Thus we must argue that representation (1) is not a good one if we assume purely accidental errors; in fact, the probability of obtaining a distribution such as the one we found for the deviations of the experimental values from a linear law (1) is exceedingly small. However, it is well known that a small value of P does not necessarily indicate that the experimental deviations from law (1) are systematic, because the "goodness of fit" test has meaning only in the case of purely accidental errors.

We believe that the very small value obtained by us for P gives good evidence for a dependence of the nuclear radius on the number of particles in a nucleus, which is more complex than the simple law (1).

[1] E. Amaldi, D. Bocciarelli, B. N. Cacciapuoti, and G. C. Trabacchi, Nuovo. Cimento, in press. Presented at the conference on elementary particles and low temperature at Cambridge, England, July, 1946.
[2] H. A. Bethe, Phys. Rev. **54**, 436 (1938); E. P. Wigner, Phys. Rev. **56**, 519 (1939).
[3] K. Pearson, *Tables for Statisticians and Biometricians*, Part I.

Search for Element 87

HORIA HULUBEI
Bucarest, Rumania
March 31, 1947

THE article by F. R. Hirsh, Jr.[1] "The Search for Element 87" came to my notice only recently because of the interruption of postal relations between America and Rumania during the war. I wish to comment on the numerical data employed in his research and on the research itself. He adopts 0.812A and 1.045A for the presumptive wave-lengths of the absorption line L_{III} and the emission line $L\alpha$, respectively. The former departs by 0.011A from the value deduced, by Moseley's law, from the most reliable experimental values obtained for neighboring elements. The latter departs by 0.017A from the value to be expected for $L\alpha_1$, and by 0.011A from the average of the to-be-expected values of $L\alpha_1$ and $L\alpha_2$. Convincing evidence for the element 87 could not be found, even assuming the element to be present in the sample, with such expectations as to the wave-lengths. I suggest that his categorical objections to the inferences of other authors are based on very inexact estimates of the emission-lines which are to be expected.

I find the experiments of Hirsh far from convincing. He used a tungsten filament and two molybdenum anticathodes, and had to clean the anticathodes frequently to remove the pulverized tungsten from them. A photograph taken with an exposure of 36 hours, and having given no indication whatever of the tungsten L lines under the conditions of his observations, should not be expected to reveal the presence of an element present very probably in much smaller quantities than the tungsten remaining on the anticathode even after hourly cleanings. In addition, the $L\gamma$ lines of tungsten would mask the presumptive line $L\alpha_1$ of element 87 if they were both emitted. Similarly, the lines of molybdenum should have made their appearance: whether or not they were sufficiently excited by the narrow residual band of continuous x-rays springing from the primary anticathode under a 25-kv bombardment and falling

on the secondary anticathode, they should have appeared as scattered radiation. What is one to think of the scattering with change of frequency considered by Hirsh and apparently the object of an inexact formula in his text? Myself, I should not have melted the sample even "gently" in order to insert it into the hole in the anticathode, in view of the presumptive properties of element 87.

The search for a stable isotope of element 87 in a mixture of heavy, volatile, and absorbing alkaline substances is difficult, even if the element should be present in proportions of 10^{-4} or 10^{-5}. The methods of spectrum analysis and the conditions of excitation must be the best attainable. Our observations were generally made by primary excitation, with tubes constructed specially for great stability of operation even when volatile substances are present. My spectra do not allow me to come to the same conclusion as that reached by Dr. Hirsh. For the reasons expounded above, I feel that his investigation does not help in the least to settle the question of the existence of a durable isotope of element 87.

[1] F. R. Hirsh, Jr., Phys. Rev. 63, 93 (1943).

The Magnetic Quenching of Superconductivity

J. W. Stout
Institute for the Study of Metals, the University of Chicago, Chicago, Illinois
April 12, 1947

In a letter with the above title Sienko and Ogg[1] suggest that for the "soft" superconductors Pb, Hg, Sn, In, Tl, CuS, Au_2Bi, Zn, and Cd the threshold magnetic field, H_T, for the destruction of superconductivity, is better represented by the expression $H_T=A(T_c^4-T^4)$ than by the parabolic relation $H_T=B(T_c^2-T^2)$. T_c is the temperature at which the metal becomes superconducting in zero field. Careful measurements of the threshold magnetic fields have been made on Sn[2,4,5], Hg[3,4], In[4,5], Tl[4,5] and Pb[4,5]. Within the accuracy of the measurements neither the parabolic nor the $\frac{4}{3}$-power law exactly fits the data. However, the parabolic law is in all cases the better approximation. The variation in the constants A ($\frac{4}{3}$-power law) and B (parabolic law) between about 1°K, and the transition temperature are: Sn (A 30 percent; B 10 percent), Hg (A 18 percent; B 3 percent), In (A 12 percent; B 8 percent), Tl (A 9 percent; B 5 percent), Pb (A 19 percent; B 8 percent). The values quoted are the averages obtained by the various experimenters. In general the variations among the measurements from different laboratories are less than the deviation from either the parabolic or $\frac{4}{3}$-power relation.

It is true that the constant A is more nearly the same for different elements than is B, but A as determined from the initial slope of H_T vs. T varies from 90 for Cd[6] to 46 for Sn, a variation well outside the experimental error and sufficient to make it doubtful that A is a combination of universal constants as Sienko and Ogg suggest.

From an expression for the threshold curve and the well-known thermodynamic relation

$$C_s-C_n=(VT/4\pi)\{(dH_T/dT)^2+H_T(d^2H_T/dT^2)\},$$

one can calculate the difference in heat capacity, C_s-C_n, between the superconducting and normal states. V is the volume. For the $\frac{4}{3}$-power relation one obtains

$$C_s-C_n=(VA^2/4\pi)(3T^3-\tfrac{4}{3}T_c^{\frac{4}{3}}T^{\frac{4}{3}}),$$

and for the parabolic relation

$$C_s-C_n=(VB^2/2\pi)(3T^3-T_c^2T).$$

Measurements on Sn[7] and Tl[8] indicate that the heat capacity of these two metals in the superconducting state obeys a T^3 law. For metals in the non-superconducting state the experimental heat capacities at low temperatures are in agreement with the theoretical prediction of an electronic heat capacity proportional to the temperature plus a lattice heat capacity which in simple cases varies as T^3. The parabolic relation predicts a difference in heat capacity of superconducting and normal metal in agreement with these considerations. The absence of a term linear in temperature from C_s-C_n obtained from the $\frac{4}{3}$-power law is inconsistent with the experimental and theoretical results for the heat capacity of metals in the non-superconducting state.

[1] M. J. Sienko and R. A. Ogg, Phys. Rev. 71, 319 (1947).
[2] W. J. de Haas and A. D. Engelkes, Physica 4, 325 (1937).
[3] A. D. Misener, Proc. Roy. Soc. A174, 262 (1940).
[4] J. G. Daunt and K. Mendelssohn, Proc. Roy. Soc. A160, 127 (1937).
[5] J. G. Daunt, A. Horseman, and K. Mendelssohn, Phil. Mag. 27, 754 (1939).
[6] N. Kurti and F. Simon, Proc. Roy. Soc. A151, 610 (1935).
[7] W. H. Keesom and P. H. Van Laer, Physica 5, 193 (1938).
[8] W. H. Keesom and J. A. Kok, Physica 1, 175 (1933).

Bounded Linear Harmonic Oscillator and Phase Transitions of Second Order

B. Suryan
Bangalore City, South India
March 28, 1947

In a recent letter in the Physical Review, Corson and Kaplan[1] have put forth the idea that the concept of a linear bounded harmonic oscillator in the theory of solids may be of help in understanding anomalies in the specific heat-curve of a crystalline solid. Similar ideas were developed by the author in 1945 and were reported in a lecture at Central College, Bangalore City, South India. These ideas were not published at the time, but in view of the considerations of Corson and Kaplan, it has seemed desirable to place them on record.

The problem arose in connection with specific heat anomalies at low temperatures in crystalline solids. It was thought that the concept of the linear bounded harmonic oscillator might be of service in studying these phenomena. The relevant theory has been worked out by Auluck and Kothari.[2] The allowed energy values are no longer $(n+\frac{1}{2})h\nu$, where ν is the frequency of the oscillator, but increase rapidly with decreasing amplitudes of oscillation. However, the present author applied the above concepts to the second-order phase transition or λ-anomalies of Ehrenfest.[3] It is well known that such phase transitions are all accompanied by marked changes in the expansion coefficient of the solid. Therefore, if at the λ-temperature the different oscillators in the crystal are bounded to different extents, the oscillators will be distributed among two well-defined

sets of energy states, and these two sets may be thought of as forming the "phases" above mentioned. This may conveniently be taken as a picture of the second-order phase transition, i.e., the crystal oscillators going from one state of bounding to another. Also because the quantum of transition between two states of the oscillator is no longer $h\nu$, a shift or a new line may be possible in the Raman spectrum or infra-red absorption of the *crystal* near the λ-temperature. As there are many substances whose λ-temperatures are near room temperature, the investigation of the Raman spectrum may provide a test of the above ideas. Also, one must consider the current theory that the specific heat change may arise from molecular rotation within the crystal. Experiments on the Raman or infra-red absorption spectrum may lead to a decision between the two conceptions.

[1] E. M. Corson and I. Kaplan, Phys. Rev. 71, 130 (1947).
[2] F. S. Auluck and D. S. Kothari, Proc. Camb. Phil. Soc. 41, 175 (1945).
[3] P. Ehrenfest, Proc. K. Acad. Weten. Amsterdam 36, 153 (1933); Comm. No. 75b from the K. Onnes Laboratory, Leyden.

On the Distribution of Electric Charge in Nuclei

B. Davison and W. H. Watson
National Research Council, Chalk River, Ontario, Canada
April 21, 1947

In a recent letter[1] by J. G. Winans entitled "A Classical Model for the Nucleus," attention was drawn to the striking relation between the atomic numbers (Z) and corresponding masses (A) of stable nuclei. Winans suggested a crystalline model in which the protons occupy surface positions as an obvious way of explaining the simple relation between Z and A.

The existence of this relation and its possible explanation by surface concentration of protons was already known to us, and we had gone so far as to test the hypothesis that protons are concentrated on or near the surface in the following simple way. In fission, if protons and neutrons are uniformly mixed the (charge/mass) ratios of the fragments should be roughly equal; whereas we should expect a higher relative concentration of charge on the light fragments, if the protons are concentrated near the surface of the nucleus before fission. We have investigated this on the basis of Coryell's tables[2] and have found that the ratio (charge/mass) is very closely the same for the light and heavy fragments. This fact, therefore, coupled with the hydrodynamical explanation[3] of the asymmetry of fission in analogy with the splitting of a classical liquid drop, seems to invalidate the idea that the protons are concentrated on the surface. In principle it is possible that some mechanism might exist to ensure constancy of Z/A for the fragments, but we know of no other nuclear phenomenon which would require such a mechanism.

[1] J. G. Winans, Phys. Rev. 71, 379 (1947).
[2] J. A. Chem. Soc. 68, 2411 (1946).
[3] J. A. Wheeler, unpublished lecture notes, August, 1946.

A Note on the Long-Lived Radio-Iodine

L. E. Glendenin and R. R. Edwards
Massachusetts Institute of Technology, Cambridge, Massachusetts*
April 9, 1947

The radioactive iodine isotope of half-life about 56 days, reported by Reid and Keston,[1] has been observed in an iodine sample separated from deuteron-bombarded Te in April, 1946, furnished by the M.I.T. Radioactivity Center.

Chemical identity of the activity was established by a series of extractions with iodine carrier. Iodide in aqueous solution was oxidized with nitrite, and the iodine extracted into CCl_4; re-extraction to water was performed with bisulfite as reducing agent. Through three such cycles, the specific activity as AgI remained constant within experimental error.

By use of a Kr-filled Geiger counter, critical absorption measurements of the x-radiations were made in Ag, Cd, In, and Sn foils. The energy of the radiations was observed to lie between the K-edges of Cd and In, characteristic of tellurium $K\alpha$ x-rays. Absorption in Al showed a small amount of an extremely soft component, and a half-thickness of 485 mg/cm^2 for the major fraction of the radiations. This corresponds to an x-ray energy of 27.5 kev, within experimental error of that expected for $K\alpha$ x-rays of Te. The soft component showed a half-thickness of 1.9 mg/cm^2 in Al, and 19 mg/cm^2 in C (polystyrene). The corresponding x-ray energies are 3.8 and 3.7 kev, respectively. The $L\alpha$ x-ray energy for Te is 3.8 kev. No particle radiations were observable from 20 mg samples (AgI or NaI) with a thin window counter, and no γ-rays of energy greater than 30 kev were apparent in the Al absorption curve. The mode of decay of the 56d I, therefore, appears to be pure orbital electron capture, with no accompanying particle or γ-radiations.

It can be shown by various considerations that the mass number of the long-lived iodine is more probably 125 than 129 as proposed by Reid and Keston.[1] The formation of I^{125} directly in the Te targets at the cyclotron is to be expected from the d, $2n$ reaction on Te^{125} (6 percent abundance) and the (d, n) reaction on Te^{124} (4.5 percent abundance). Considerations which render improbable the mass assignment of 129 are: (a) the proposed decay scheme[1] requires either that Te^{129} have a stable isomer, or (for the chain to reach a stable end product) that a unique chain of K-capture followed by two β^- decays occur (i.e., I^{129}—K→ Te^{129}—β^-→I^{129}—β^-→Xe^{129} stable); (b) according to Bohr and Wheeler[2] the most stable nuclear charge (Z_A) for mass 129 is near 54 (Xe); thus I^{129} should have no appreciable probability for decay by K-capture, but should decay by β^- emission to stable Xe^{129}; (c) no radio-iodine of half-life near 56 days is reported among the U^{235} fission products,[3] whereas I^{129} must certainly be formed; the latter is listed as a very long-lived, hypothetical descendant of 70m Te^{129}.

The nucleus I^{125}, however, would certainly decay by K-capture or positron emission to stable Te^{125} and would not be formed in appreciable yield as a U^{235} fission product since it is shielded by stable Te^{125} from formation by β^- decay. Its predicted decay energy for K-capture, calculated from

the equation of Bohr and Wheeler,[2] is about 0.2 Mev. This is concordant with the fact that no energetic radiations are observed.

Possible mass assignments[4] other than 125 may be tentatively ruled out. On the basis of the Bohr-Wheeler[2] picture, I^{128} and isomeric states of the nuclei of mass 124, 126, 128, and 130 would be expected to have high decay energies (>1 Mev). A radioactive isomer of stable I^{127} would decay by isomeric transition, and thus would yield no Te x-rays. Iodine isotopes of mass number greater than 130 have been conclusively characterized as energetic β^- emitters.[3]

The yield of the 56*d* I in deuteron-bombarded Te targets at the M.I.T. cyclotron is about $15\mu c/\mu$ amp. hr. of 14-Mev deuterons, calculated on an assumption of one *K* x-ray per disintegration, and an experimentally determined counting efficiency of about 1 percent in the usual argon-filled Geiger counter. The corresponding yield of 12h I^{130} is 750 $\mu c/\mu$ amp. hr. (for short bombardments). The yield ratio of 12h I to 56d I at saturation is 4.5. The abundance ratio of Te^{130} to Te^{125}, from which I^{130} and I^{125} are formed by the $(d, 2n)$ reaction, is 5.5. Assuming the (d, n) reaction on Te^{124} to be less probable than $(d, 2n)$ on Te^{125}, the proposed mass assignment of 125 for the long-lived I is consistent with the observed yield.

We wish to thank Professor M. Deutsch of the Physics Department and Professors C. D. Coryell and J. W. Irvine, Jr., of the Chemistry Department for valuable suggestions and criticisms, and the Radioactivity Center for supplying the sample of radioactive iodine.

* Contribution from the Department of Chemistry and the Laboratory for Nuclear Science and Engineering; supported in part by U. S. Navy Department, contract N5ori78 Task VI.

[1] A. F. Reid and A. S. Keston, Phys. Rev. **70**, 987 (1946).
[2] N. Bohr and J. A. Wheeler, Phys. Rev. **56**, 426 (1939).
[3] "Nuclei Formed in Fission," J. Am. Chem. Soc. **68**, 2411 (1946).
[4] G. T. Seaborg, Rev. Mod. Phys. **16**, 1 (1944).

55-Hour Element 61 Formed in Fission*

MARK G. INGHRAM, DAVID C. HESS, JR., AND RICHARD J. HAYDEN, *Argonne National Laboratory, Chicago, Illinois*

AND

GEORGE W. PARKER, *Monsanto Chemical Company, Clinton Laboratory, Oak Ridge, Tennessee*

April 15, 1947

A CHEMICAL separation of element 61 from fission products was made to verify the mass assignment and to study the activity of 49-hr. 61^{149}, as summarized in fission tables.[1] This activity has been assumed to have a mass of 149 from fission yield data.

An aliquot of the element 61 sample was placed on the filament of a mass spectrograph, and normal neodymium was added to serve as a mass standard. By operation of the spectrograph the isotopes were separated according to mass and deposited on a photographic plate. A second "transfer" photographic plate was placed face to face with the first plate, and four developable images were obtained from its radioactive isotopes which by comparison with the normal neodymium spectrum on the first plate were shown to have masses 147, 149, 163, and 165. The metal lines at 147 and 149 were very weak in comparison with the metal oxide lines at 163 and 165 but appeared in approximately the same ratio. Since the ratios of metal to metal oxide are different for the different elements which might have given these masses, this indicates that all lines are due to the same element.

To verify that the activity at the mass 149 position was 49-hr. 61^{149} the following technique was used. The photographic plate containing the active isotopes was placed successively against various parts of a larger photographic plate for times calculated to give equal intensity at the mass 149 position if the half-life were 49 hours. Five exposures were taken for successive times of 13.4 hours, 16.5 hours, 21.6 hours, 31.0 hours, and 57.3 hours. These transfers showed the 61^{149} to be decaying with a half-life of slightly more than 49 hours, while the 147 showed no detectable decay. A decay curve taken simultaneously on another aliquot of the element 61 sample showed it to contain a 55-hour activity and a long-lived activity. Thus the 147 and 163 lines are due to the metal and metal oxide ions of 3.7-year 61^{147} as shown previously by Hayden,[2] and the lines at 149 and 165 are due to the metal and metal oxide ions of 55-hr. 61^{149}.

The samarium-europium fraction of the original fission sample was also analyzed by this method. It showed activities at masses 151, 153, 155, 156, 167, and 169. The active lines at mass numbers 153 and 169 were found to decay with a half-life of 47 hours. Since the ratio of intensities of the 153 to 169 is that characteristic of Sm^+ and SmO^+ this activity is 47-hr. Sm^{153} as was shown previously by neutron irradiation of samarium.[3] The lines at 151 and 167 were also ascribed to samarium since the ratio of the intensities was the same as the 153 and 169. This is the activity first found in fission by Lewis and Hayden. From intensity considerations the half-life of this activity has been estimated at roughly 20 years. The lines at 155 and 156 were assigned to europium since only the metal ion lines appeared. The 156 decayed with a half-life of about 15.4 days and the 155 showed no detectable decay. Thus these activities are due to 2 to 3 year Eu^{155} and 15.4-day Eu^{156}.

* This document is based on work performed under Contract No. W-7401-eng-37 for the Manhattan Project at the Argonne National Laboratory.

[1] Rev. Mod. Phys. **18**, 513 (1946).
[2] Richard J. Hayden, Phys. Rev. (1947), to be published.
[3] Richard J. Hayden and Mark G. Inghram, Phys. Rev. **70**, 89 (1946).

Proceedings of the American Physical Society

Minutes of the Meeting of the New York State Section at Rochester University, April 12, 1947

THE Spring Meeting of the New York State Section of the American Physical Society was held at Rochester University on Saturday, April 12, 1947. An attendance of over 130 members and visitors was recorded.

The following program of invited papers was presented:

Nuclear Physics and Machines. S. N. Van Voorhis, *University of Rochester.*

The 130 Cyclotron at Rochester. S. W. Barnes, *University of Rochester.*

Photographic Plates for Use in Nuclear Physics. J. H. Webb, *Eastman Kodak Company.*

Separation of Stable Tracer Isotopes. D. W. Stewart, *Eastman Kodak Company.*

Biological Effects of Nuclear Radiation. W. F. Bale, *University of Rochester.*

Plans of the Brookhaven National Laboratory. G. B. Collins, *University of Rochester.*

The Association of Rochester Scientists, The Executive Committee Rochester Association.

C. L. Andrews,
Secretary